CALIBRATION OF THE SOFTWARE ARCHITECTURE SIZING AND ESTIMATION TOOL (SASET)

THESIS

Carl D. Vegas, B.S.
1st Lieutenant, USAF

AFIT/GCA/LAS/95S-11

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY
AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT/GCA/LAS/95S-11

19951117 016

CALIBRATION OF THE SOFTWARE ARCHITECTURE
SIZING AND ESTIMATION TOOL (SASET)

THESIS

Carl D. Vegas, B.S.
1st Lieutenant, USAF

AFIT/GCA/LAS/95S-11

Approved for public release; distribution unlimited

AFIT/GCA/LAS/95S-11

CALIBRATION OF THE SOFTWARE ARCHITECTURE SIZING AND ESTIMATION TOOL (SASET)

THESIS

Presented to the Faculty of the Graduate School of Logistics

and Acquisition Management of the Air Force Institute of Technology

Air University

In Partial Fulfillment of the

Requirements for the Degree of

Master of Science in Cost Analysis

Carl D. Vegas, B.S.

1st Lieutenant, USAF

September 1995

Approved for public release; distribution unlimited

Acknowledgments

I would like to take a moment to thank the people who helped make this research a reality. Without their enthusiastic support this would undoubtedly have been a much more painful experience.

"Thanks" go out to Mrs. Sherry Stukes and Mrs. Gina Novak-Ley for being such supportive sponsors by providing the data and useful feedback throughout the entire process. Their frequent trips for face-to-face communication made a big difference. I would also like to thank Mr. Tom Pighetti and Mr. Rick Maness of Lockheed-Martin for their friendly assistance in my endeavor to understand the finer points of SASET.

I thank Professor Dan Ferens for being such an open-minded and helpful thesis advisor. His efforts to provide timely feedback and to be available to help me overcome the stumbling blocks are greatly appreciated. Thanks to Dr. Dave Christensen for providing his valuable opinions as my reader.

Lastly, I would like to acknowledge the support and encouragement provided by my family. Thanks for always helping me regain my balance and climb back into the ring.

Dave Vegas

Table of Contents

List of Figures

List of Tables

List of Equations

AFIT/GCA/LAS/95S-11

Abstract

This study attempted to analyze the effect of calibration on the performance of the SASET computer software cost estimating model. Data used for input into the model were drawn from the most current USAF SMC Software Database (SWDB). Once all the records to be used for analysis were identified, the DBMS/Calibration tool (which is part of SASET) was used to perform regression analysis on the relationship between program size (measured in SLOC) and the effort required to develop the program (measured in man-months). Productivity information reported from this tool was then input into equations used to calculate the Productivity Calibration Constants (PCC) and Software Class Multipliers. A comparison was then made between the model's accuracy before calibration and its accuracy after calibration. This was done using records which were not used in calibration (referred to as validation points). Several measures such as mean, variance, mean magnitude of relative error (MMRE), and the percentage method were used to describe accuracy. The majority of the results agreed with previous studies that calibration does improve a model's prediction performance. However, emphasis is placed on the fact that calibration is most useful when the group of calibration data points are homogenous.

CALIBRATION OF THE SOFTWARE ARCHITECTURE SIZING AND ESTIMATION TOOL (SASET)

I. INTRODUCTION

General Issue

Computers have an increasingly vital role in our personal and professional lives. They help us do many things more efficiently and effectively. They have greatly increased the rate at which information is transferred by reducing the constraining effects of time and distance. In the Department of Defense (DoD), they help us do everything from writing an evaluation on a standardized form to placing a bomb within inches of its target. The improvements in technology which make our tasks easier are the result of improvements in both hardware *and* software. However, the increasing usefulness of software in various applications is making its costs an increasingly greater proportion of the total cost of computer systems (Boehm, 1984). In fact, the DoD alone currently spends $30 billion annually on software (Ferens, 1994: 1). The tremendous growth of investment in software can be seen by comparing this statistic to the fact that in 1980 the US *as a whole* spent $40 billion on software (Boehm, 1984). As a result, software has become very "high visibility" and estimates of its acquisition and maintenance costs in future projects are of

great concern to the DoD. While the maintenance or life-cycle costs of software are important, this research effort will limit its focus to acquisition cost estimation.

Unfortunately, current software cost estimating models have not shown significant increases in accuracy over models decades old (Boehm, 1984). This is the reason more effort needs to be put into developing more accurate models and improving the performance of current models through calibration; which is defined in the "Definitions" section later in this chapter. In fact, Thibodeau stated in his report, "we have shown that the calibration of model parameters may be as important as model structure in explaining estimating accuracy" (1981: 6-6).

A major reason for this is the change in the factors affecting cost. As the processes and products used to develop software improve, the validity and relevance of historical data points may consequently be diminished. Recurring calibration based on more recent projects can help overcome this problem.

Of course, another obvious reason that a model's estimates may be inaccurate is the possibility of inaccuracy in the user-estimated inputs. Improvement in this area is not related to the model but instead depends upon the user's ability to predict the resources required by a future project.

One model that has been developed by the Martin-Marietta Corporation for the Naval Center for Cost Analysis (NCA) and which is being maintained by the Air Force Cost Analysis Agency (AFCAA) is the Software Architecture Sizing and Estimation Tool (SASET) model (Bowden, Cheadle, & Ratliff, 1993b). Just like other cost estimating models, SASET is comprised of equations which require the user to input values for

certain parameters (substitutes for cost drivers) to come up with an estimate. The parameters used in a given model vary and are those determined to be important by the creator of the model (in this case the Martin Marietta Corporation; now Lockheed-Martin). The essential differences between the various models is that each uses different cost estimating relationships and each emphasizes different parameters within those relationships.

Specific Issue

The Task. The Air Force Space and Missile Systems Center (SMC) is one of the many DoD organizations that invests heavily in computer systems. This is due to the "high tech" nature of the space systems they manage. The precision and life-critical requirements of such systems lead to the need for highly reliable software. "High reliability" requires a great deal of testing and, consequently, a great deal of money. SMC has therefore expressed an interest in having SASET calibrated to a database they possess which contains historical software data for previous space and missile projects (a.k.a. unmanned space projects); as well as several other types of projects. Specifically, the "other" types of projects to be analyzed are: ground, missile, mobile, avionics, and commercial.

The primary purpose of this research effort is to aid DoD decision makers by providing a calibration factor (based on the most current data available) that improves SASET's accuracy in predicting future unmanned space project software effort (cost). Secondary goals of this effort are to provide calibration factors for the other types of

projects contained in the database and to provide the reader with a step-by-step reference of how to calibrate the current version of SASET to their own database.

Calibration will be achieved by regressing a project's size against the effort (or manhours) required to create the project. This will be done for selected records from the database; the selection process will be described throughout Chapters 3 and 4. Calibration will be performed with the DBMS/Calibration Tool; which is part of SASET. From this regression, the DBMS will provide a "productivity" value which reflects the amount of labor hours required per lines of code produced (Harbert, 1993: 4-5). Dividing this value by the "productivity reference" value in the calibration file of the DBMS (the average productivity value for ground - application projects) will give the "software class" effort multiplier. Since SASET uses ground software as a reference point, it initially creates an estimate for the parameters entered by the user as if the project were to be ground software. This is the reason it is necessary to adjust for different software classes by using multipliers. These multipliers are placed into the SASET calibration file and automatically increase or decrease the estimate produced by SASET to adjust for the relative difficulty (or ease) of a software class as compared to a ground project with all the same inputs. Additionally, a separate multiplier is used to adjust for the "type" of software the project represents. Definitions of these terms are found later in this chapter and the specific steps which must be performed will be discussed in Chapters 3 and 4.

"Accuracy" will be determined by statistical testing of the calibrated model's estimates in comparison to the actual costs. More important than overall accuracy (for the purposes of this research) is the relative *increase* in accuracy, if any, brought about by

calibration. Various measures will be employed to detect increased accuracy. These measures are specified in Chapter 3.

SASET: The model. Descriptions of SASET can be found in various sources: the SASET user's guide, the "Air Force Cost Analysis Agency Software Model Content Study" done by Management Consulting & Research Inc. (MCR) in 1994 (Stukes, 1994a), and the Coggins & Russell Air Force Institute of Technology thesis in 1993 (Coggins, 1993). It may be more useful to refer to the independent studies (the last two listed above) if one is attempting to identify the model's strengths and weaknesses relative to other software cost estimating models in use throughout the Air Force.

Here is some basic information about the model. The current version is 3.0 and was released April 1993. It was developed by Martin Marietta Astronautics Group, was sponsored by the Naval Center for Cost Analysis, and is maintained by the Air Force Cost Analysis Agency. It is distributed through the AFCAA and is available, by request, to commercial organizations. It is PC-based and is not password protected. A more detailed description is provided in Chapter 2.

Calibration. Calibration is the adjustment of a model's equations to induce the model to provide a predicted outcome as close as possible to the actual outcome for a given set of data. There have been previous attempts at calibrating SASET; two were done in 1991. One of them was performed as an AFIT thesis by Capt. Gerald Ourada. Unfortunately, he stated that "since . . . a calibration mode is not available as part of the computerized version of the model, [he] could not figure out how to calibrate the model

to a particular data set" (1991: 4.8). He was working with the older version (2.0), which had a calibration file but no calibration tool.

The other attempt was made by MCR for the Space Systems Division or SSD (now referred to as the Space and Missile Systems Center). Apparently they were able to get an early release of the current version which, unlike its predecessor, supports recalibration via its Database Management System (DBMS) and Calibration Tool. This tool allows the user to perform various types of regressions, the result of which are Productivity Calibration Constants (PCC) for the three "types" of software recognized by SASET: system, application, and support. The PCC for each class of software in the SSD Software Development Database (SDDB) used by MCR are given in their report. The SDDB happens to be an older version (with fewer records) of the database being used for the current research effort (the SWDB). However, the change in predictive accuracy of the model after calibration is not contained in the MCR report. The DBMS is discussed in more detail in the next chapter.

Research Objective

In order to effectively calibrate SASET, I will have to address the following basic questions (for each class of software):

1. What is SASET's pre-calibration accuracy with the data set selected for validation?

2. How is SASET calibrated?

3. After calibrating SASET with the selected data set, what is the model's accuracy with the validation data set?

4. What is the improvement in accuracy after calibration?

Scope Of Research

The scope of this research effort is limited to calibration parameters derived for the operational environment reflected by the SMC database described below. Determination of the usability of the calibrated model from this research across different environments could be an area for future research utilizing appropriate databases. No inferences will be made as to the ability to calibrate any other model to this database. Also, as mentioned previously, this analysis will be limited to development effort only.

Definitions

The following are some useful definitions for understanding the results of this research:

Calibration - adjustment of the model equations to induce the model to provide a predicted outcome as close as possible to the actual outcome for a given set of data

Class - a classification of software which identifies the physical environment in which the software will operate or the system in which it will be employed (i.e. space, ground, missile)

DBMS - a calibration and database utility used primarily to recalibrate SASET

Effort- the number of manhours or manmonths required to produce SLOC

Manmonths- a measurement unit of the effort required to produce a software program; the standard is 152 hours of labor per manmonth

PCC - "productivity calibration constants" adjust SASET estimates for the differences in the difficulty of producing software of different types of software

SASET- "Software Architecture Sizing and Estimating Tool"; a parametric software cost estimating model

SLOC- "source lines of code" is a measurement unit of the size of a software program

SMC- Space and Missile Systems Center

Software Class Effort Multiplier- adjusts for the level of effort typically required by a given class as compared to ground software projects

Stratify- to divide data into homogenous groups in order to perform analysis and discover patterns; more detailed subdivisions usually reduce the number of useable points.

SWDB- "Software Database" is the database created by SMC and used for this research

Tier- a term used in SASET to refer to the different levels of information or inputs used by the model; the five tiers are listed in Chapter 2

Type - this term is used in several ways throughout this document

-- Software type: is a classification of software which is related to the function the software will perform; the three categories in SASET are application, system, and support

-- Project type: synonymous with the "class" of the project

-- Regression type: the regression equations specified in the DBMS

Validation - process of determining the accuracy of the model; the difference between the model's predicted outcome and the actual outcome for a set of data similar, but not identical, to the set used in calibration.

The next chapter will provide the interested reader with a more general discussion on cost estimation and historical calibration research.

II. LITERATURE REVIEW

Overview

This chapter is a review of relevant material and ideas and is intended to provide the reader with a brief, but useful, synopsis of the basics of software cost estimation and the current status of the software cost estimation field. Also found in this chapter is a more detailed discussion of SASET and DBMS.

Software Estimation

Just like any other endeavor, software engineering is limited by the resources available (Boehm, 1984: 239). Trade-offs must be made between cost, time, hardware capacity, and personnel skill and availability. Therefore, the ability to estimate a total cost is desirable since it can help guide such decisions during the planning stages. This is the idea behind software cost estimation.

One might be tempted to say, "the average cost of this type of software is ...." However, due to the complex nature of software development and all the factors that drive the costs of a project, it is likely that a better estimate can be achieved by "modeling" software development rather than just using "averages." A mathematical model is built to simulate reality as closely as possible. It can never be perfect because there are an infinite number of factors that play a role in driving costs. The key is to capture all or most of the *significant* factors.

There are seven major software cost estimation techniques (which are not mutually exclusive):

> **1) algorithmic (or parametric) models**: these methods provide one or more algorithms which produce a software cost estimate as a function of a number of variables which are considered to be the major cost drivers [i.e. $Y = Ax^B$].
>
> **2) expert judgment**: this method involves consulting one or more experts, perhaps with the aid of an expert-consensus mechanism such as the Delphi technique.
>
> **3) analogy**: this method involves reasoning by analogy with one or more completed projects to relate their actual costs to an estimate of the cost of a similar new project.
>
> **4) Parkinson**: a Parkinson principle ('work expands to fill the available volume') is invoked to equate the cost estimate to the available resources.
>
> **5) price-to-win**: here, the cost estimate is equated to the price believed necessary to win the job (or the schedule believed necessary to be first in the market with a new product, etc.).
>
> **6) top-down**: an overall cost estimate for the project is derived from global properties of the software product. The total cost is then split up among the various components.
>
> **7) bottom-up**: each component of the software job is separately estimated, and the results aggregated to produce an estimate for the overall job. (Boehm, 1984: 242)

According to Dr. Boehm's article, two of these techniques are "unacceptable and do not produce satisfactory estimates" (Boehm, 1984: 242): Parkinson and price-to-win. He also states that each of the other techniques has its own unique strengths and weaknesses. It is also possible that there may be some overlap between these techniques. The algorithmic/parametric technique derives its name from the fact that it is a mathematical model comprised of various parameters (or cost drivers) to which the user

can assign values. Thus the algorithmic/parametric technique can be said to be quantitative *and* qualitative; it assigns values to certain parameters (values estimated by the user) to mathematically represent real-life scenarios of resource quality and availability.

The strengths associated with this technique are: "objective, repeatable, analyzable formula; efficient, good for sensitivity analysis; objectively calibrated to experience" and the weaknesses are "subjective inputs; assessment of exceptional circumstances; calibrated to past not future" (Boehm, 1984: 243). Yet it should be noted that the strengths may also become weaknesses if the input is invalid; a classic "garbage in, garbage out (GIGO)" system.

Professor Ferens states that most of the software cost models we see in the DoD are a combination of the algorithmic and top-down approaches. Some other strengths and weaknesses are offered by Professor Ferens for this "combination" are: "easy to use; fast; useful early; reliability" but "inaccurate; unstable (some sensitive drivers)" (Ferens, 1994).

Due to the nature of cost estimation, all of the techniques can be said to have some inputs that are inherently more subjective or judgmental than other inputs. For example, the quality of the programmers is hard to quantify. Since we are attempting to predict as far into the future as possible it is unavoidable that we have some subjective inputs. Also, each model represents the ideas of its developer as to the fundamental relationships between various cost drivers (represented by input parameters) and cost. Thus, each model emphasizes different drivers. These differences are what cause different models to provide different estimates for the same project. Yet each of the models is said to be "reliable" since it will always give you a certain output if you enter the same inputs. It

should be mentioned that one important similarity between the majority of models in use is that they emphasize the importance of size (SLOC) as a cost driver.

If the user-forecasted inputs turn out to be correct, the outputs will be correct *if the model accurately represents reality.* "Accuracy" is, in fact, the major problem with software cost estimating models today. In 1984, Dr. Boehm noted that the state of the art in software cost models at that time was estimates within 20% of actuals about 70% of the time (Boehm, 1984: 251). Not much improvement has occurred since then. This may be partially due to a flawed representation of reality by the model but, as mentioned previously, changes in software development have a major impact on a model's ability to make predictions based on historical data. In fact, some argue that reported inaccuracies of model estimates are *primarily* due to the rapid changes in development tools and procedures constantly experienced in this dynamic field.

This author is aware of one study that reported the change in SASET's estimation accuracy due to calibration; although the actual objective of the study was to determine if there was a need for an Ada-specific estimating model. The study was done in 1989 by the Illinois Institute of Technology Research Institute (IITRI) and looked at both effort and schedule estimation. It compared the estimates of six different software cost estimating models to the actual effort and schedule data of eight different Ada language software development projects.

After normalization of the models' outputs (for SASET, this consisted of subtracting the Quality Assurance effort estimate from the total effort estimate), they found that SASET's overall accuracy for predicting effort for these eight projects was that

it estimated four out of the eight projects to within $\pm$ 29% (IITRI, 1989: 3-14). The term the study used to identify the calibrated model results was "overall consistency"; since they were trying to eliminate user input bias by determining whether a given model's estimates were consistently high or consistently low (IITRI, 1989: 3-17). The actual calibration was performed through the following steps (IITRI, 1989: 3-17):

1. A percentage of actual effort to model effort was calculated
2. The two extremes were discarded to achieve a truer sampling of percentages
3. A mean value of the remaining percentages was computed and applied to the given model's estimates
4. The relative error for each project was recalculated using the adjusted efforts.

While this author is not certain about the mathematical validity of this procedure, the results were as follows: SASET now had a range of -15% to 27%, 50% of the time (four out of eight projects) (IITRI, 1989: 3-19). Thus we see an improvement in performance after calibration. However, the specific results of this current research should not be rigidly compared to the IITRI study results for several reasons: they used SASET version 1.5 to make the estimates, the calibration was performed in a different manner than will be done here, different statistics will be used to measure the effects of the calibration, and this research will look at more data and not restrict itself to Ada projects. Yet, the fact that there was an improvement lends justification to this current research and encourages further investigation into the usefulness of calibration.

SASET

SASET is a model which employs the "algorithmic/parametric" technique of estimation. The model's equation for calculation of effort is (Bowden, 1993a: 2-23):

$$\text{Effort} = K * \text{class of software adjustment factor} * \text{EQType} * \text{ASBCM} \qquad (1)$$

where K = productivity calibration constant (PCC)

EQType = Equivalent HOL for each software type

ASBCM = Average software budget complexity multiplier

This research effort will analyze the effect on prediction accuracy of adjusting "K" and "class of software adjustment factor."

SASET was developed in four years (1986 through 1990) by Martin Marietta Astronautics Group under contract for the NCA. Enhancements to the model (resulting in the current version) were implemented from 1990 to 1993 under follow-on contracts with the both the NCA and the AFCAA. As stated in the User's Guide, "SASET is used by estimators, planners, software developers and managers to size, cost, and to establish schedules for software development projects" (Bowden, 1993b: 1-1).

The model is organized into five tiers:

- system environment factors (Tier 1)
- software sizing (Tier 2)
- system attributes (Tier 3)
- maintenance/support (Tier 4)
- risk assessment (Tier 5)

Only the first three tiers are required before SASET will give cost and schedule estimates for development projects. These are the only tiers that will be discussed in this research effort.

Tier 1 allows the user to specify several parameters that describe the overall project such as the "class" of software (ground, avionics, commercial, etc.), programming language (higher order language (HOL) versus assembly), development schedule, and other "developmental issues."

Tier 2 deals with the expected size of the software to be developed. Size tends to be the primary cost driver in most software cost estimating models. This tier allows the user to either directly enter the number of source lines of code (SLOC) he or she predicts or to use the "functionality" input method whereby SASET determines the expected SLOC from the user's inputs as to the number and type of functions the software is to perform. In this tier, the user can also specify the type of software (system, application, or support), the language (HOL versus assembler), and the amount of new/modified/reused code.

Finally, Tier 3 is where the user is allowed to describe unique project attributes through various parameters which SASET identifies as affecting, negatively or positively, the complexity of the project. Some examples of the attributes contained in this tier are experience of programmer personnel, documentation required, hardware availability and compatibility, and number of development sites (see Table 1 below). These attributes are set on a scale of 1 through 4, with 3 being "nominal" or average. The user must decide whether to change the rating of a given attribute within this scale and can receive guidance

for each attribute (although sometimes minimal) from on-screen help via the F1 key. The narratives found in the "help windows" are essentially the same as those found in Chapter 4 of the User's Guide.

TABLE 1. SASET Inputs

TIER	*INPUTS*
Tier 1 "System Environment"	System Environment: Class of S/W; H/W system type; % of memory utilized; # S/W configuration items; Schedule; # Development locations; # Customer locations; # Workstation types; Primary S/W language; % of micro-code; Lifecycle choice.
Tier 2 "Software Sizing"	High Order Language code: new (systems, applications, support); modified (systems, applications, support); rehosted (systems, applications, support). Assembly code: new (systems, applications, support); modified (systems, applications, support); rehosted (systems, applications, support). Data Statements.
Tier 3 "System Attributes"	System Attributes: System requirements; S/W requirements; S/W documentation; Travel requirements; Man interaction; Timing & Criticality; S/W testability; H/W constraints; H/W experience; S/W experience; S/W interfaces; Development facilities; Development vs. host system; Technology impacts; COTS S/W; Development team; Embedded development system; S/W development tools; Personnel resources; Programming language. CSCI Integration Factors: S/W language complexity; Modularity of S/W; S/W timing & criticality; # of CSCI interfaces; S/W documentation; Development facilities; S/W interfaces; Testing complexity; Development complexity; Integration experience; Integration development tools; Schedule constraints.
	NOTE: S/W = software; H/W = hardware

Perhaps one of the most notable features of the model, in light of this research effort, is that it allows the user to change virtually all of the numerous parameters in its calibration file. This may seem daunting at first but, as mentioned in Chapter 1, the model comes with a calibration tool referred to as the Data Base Management System (DBMS) to aid the user in such an effort. Note that this is an optional procedure but, as mentioned

in the previous chapter, the accuracy of a model's estimates for a given environment may be improved by calibrating the model to that environment using data that is as current as possible. This makes intuitive sense. Since parametric models base their estimates on historical cost estimating relationships, it should improve estimation accuracy to adjust the coefficients of the model's estimating equations based on the most current data available. "In general, DBMS allows collection and storage of past software projects. From these projects, regression fits are performed to derive Productivity Calibration Constants (PCCs) for software types...These PCCs are imported into the SASET calibration file" (Harbert, 1993: 1-2). Adjustment for the class of software is also necessary. The actual steps for both adjustments are described in Chapter 3.

Thus, DBMS is a sort of automatic calibration program within SASET. However, it is up to the user to provide the data required for the DBMS to run its various types of regression, to choose the type of regression desired, and to determine whether the calibrated model is of more value than the uncalibrated one. Also, as will be discussed in Chapters 3 and 4, there are quite a few manual (and undocumented) steps the user must perform to calibrate SASET. The DBMS is described in more detail in Chapter 3.

A final word of caution is in order here. One must remember that regression analysis is useless for estimation if the system to be developed is *radically* different than those which were used to come up with the regression line (or calibrated model). What constitutes "radically" must be determined by the individual doing the estimating.

III. METHODOLOGY

Overview

This chapter discusses the procedures which will be employed in this research effort and includes discussions on the type of data to be used, the tool to be used, the design to be followed, the statistical measures to be employed, and the expected results.

Data Description

The data used in this research is from the Space and Missile Systems Center (SMC) database version 1.0; referred to as the Software Database (SWDB). "The SWDB was developed under the direction of the USAF SMC, with assistance from the Space Systems Cost Analysis Group (SSCAG)" (Fulton, 1995; 2). This database contains over 2,000 data records or projects (primarily space and missile projects) and allows the user to query or sort the records in various ways. For this research effort, the database will be divided into homogenous groups in order to determine the specific data sets (or "samples") to be used for the calibration and validation of SASET. These data sets will be chosen based on several criteria:

1. the data points are from the same software class (i.e. unmanned space, avionics),
2. the data points identify as many model inputs as possible (especially key inputs for SASET such as software type and development standard),
3. there are enough points to be able to divide up between the calibration and the validation procedures (DBMS requires a minimum of five records for calibration)

4. the records contain effort and SLOC information.

DBMS

> DBMS is a calibration and database utility used primarily to recalibrate the SASET model. In general, DBMS allows collection and storage of past software projects. From these projects, regression fits are performed to derive Productivity Calibration Constants (PCC) for software types (System, Application, and Support). These PCCs are imported into the SASET calibration file. The DBMS utility may also be used in a stand-alone analysis mode. Direct keyboard access is provided for input data. The resultant output regressions and associated tables are displayed for subsequent analysis. (Harbert, 1993: 1-2)

The tool is basically broken down into two sections: data collection and calibration. The data collection section is the database where the historical data records are entered. Each record has three levels of input: project information, sizing information, and budget information. The project information is general information which allows for identification of a given record. The last two levels of input are the ones used for the regression analysis. The type, quantity, language, and condition of the source lines of code (SLOC) are contained in the sizing information window. Effort (in manmonths) is broken out by phase and organization in the budget information window. This section allows some basic database procedures to be performed; such as browsing, updating, sorting, etc.

Once this database is established and all the records to be used for calibration have been "marked," the calibration section of the tool can be used. Calibration can be based on either the "type" of software or the "complexity" of the project. It is up to the user to

decide which is chosen; this research will only look at regression by type since this is a more objective classification.

Next, the user must choose from among the various types of regression DBMS performs; effort (in manmonths) is regressed against SLOC. It is the author's opinion that the user should decide which seems logical a priori if the data set to be calibrated is small. This prevents choosing a regression type which is chasing the error rather than finding a useful estimation line. The DBMS documentation does not provide much help in this decision. The only types which the author felt to be feasible choices were the linear, power, and logarithm. The author recommends choosing from among these three regression types. More advice on how to decide on which is appropriate will be provided throughout this document. However, the user does have the freedom to choose from among the different types and is aided by the graphical and table outputs. The graphs allow for a visual interpretation of the best fit line and the "regression summary" tables for each type of regression report the root mean squared residual (RMS) value; the lower the RMS the better the fit. RMS and correlation are the only statistics reported in DBMS.

After choosing the type of regression believed to be most valid, the user then accesses the "overall summary table" to discover the "average productivity" value. This is the PCC for the given type and class of software being analyzed. It is recommended that only one regression type be highlighted at a time since interpretation of a productivity value found by simultaneously choosing more than one regression is unclear. This value is then used to adjust both the PCC and the "software class effort multiplier" in the

calibration file. The specific steps and calculations of the entire process are listed later in this chapter.

Research Design

This research essentially consists of calibration and validation. "Calibration" is the adjustment of the model equations to induce the model to provide a predicted outcome as close as possible to the actual outcome for a given set of data. "Validation" is the process of determining the accuracy of the model; the difference between the model's predicted outcome and the actual outcome for a set of data similar, but not identical, to the set used in calibration.

The first step is to stratify the data. This should be based on the way the model is structured. Due to SASET's structure, this will be done by dividing the data records into sets based upon "class" of software and then subdividing based upon "type" of software. The next step is to identify a set of data, for each class, that is useful for the analysis (has sufficient information). Part of this set is to be used for calibration and the remainder for validation. The rule of thumb, decided upon by SMC, to be used to divide the data set between calibration and validation is as follows:

- If the data points are ≤ 8, then use all the points for calibration only
- If the data points are >8 but ≤ 11, then use 8 for calibration and the rest for validation
- If the data points are ≥ 12, then use 1/3 for validation.

The data points used for validation will be assigned randomly. This is necessary to reduce the possibility of bias and thus improve the likelihood that the validation results

truly reflect the performance of the calibrated model in the given environment. This will be done by listing the candidate records in ascending order of SLOC and then selecting every third record for validation. The reason the records will be arranged by SLOC is to ensure the use of data points that represent the range of SLOC contained in the database for each software class. This should result in a more useful and representative regression line as well as representative validation points.

One would expect the validation to reveal that the accuracy of the calibrated model is superior to the uncalibrated model since, as mentioned, the validation data is taken from the same data set as the calibration subset.

To determine estimating accuracy, I will compute mean, variance, and standard deviation for the *differences* between actual and estimated effort (equations 4, 5, and 6 below). The square root of the coefficient of multiple determination (R^2) between SLOC and effort will be used to analyze selection of appropriate regression types; although its square root (the correlation coefficient) is reported by the DBMS, equations (2 and 3) below were used to verify the value reported.

I will also use three of the statistical analysis measures employed by Ourada in his calibration and validation effort of four DoD software cost estimating models (1991): the magnitude of relative error (MRE) (equation 7) and mean magnitude of relative error (MMRE) (equation 8), the prediction level test (or "percentage method") (equation 9). As used here, the percentage method describes the number of estimates (of effort) within 25% of the actual (effort) as a percent.

The Wilcoxon Signed-Rank Test (equation 10) will also be used in order to identify estimation bias. This is a nonparametric alternative to the parametric "Paired T Test." In fact, we can say "the Wilcoxon test is never very much less efficient than the *t* test and may be much more efficient if the underlying distribution is far from normal" (Devore, 1991: 608). Thus, using this test allows us to avoid making assumptions about the distributions of the "actuals" and the "estimates."

The actual equations for all the statistics mentioned above are as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (E_i - \hat{E}_i)^2}{\sum_{i=1}^{n} (E_i - \overline{E})^2} \tag{2}$$

where E = actual effort, E-hat = estimated effort, and E-bar = mean effort

$$r = \sqrt{R^2} \tag{3}$$

$$\overline{E} = \frac{1}{n} \sum_{i=1}^{n} \hat{E}_i \tag{4}$$

where n = number of records

$$s^2 = \frac{n \sum_{i=1}^{n} x_i^{\,2} - \left(\sum_{i=1}^{n} x_i \right)^2}{n^2} \tag{5}$$

where x = E - E-hat

$$s = \sqrt{s^2} \tag{6}$$

$$MRE = \left| \frac{E - \hat{E}}{E} \right| \tag{7}$$

$$MMRE = \frac{1}{n} \sum_{i=1}^{n} MRE_i \tag{8}$$

$$PRED(l) = \frac{k}{n} \tag{9}$$

where k = number of estimates within 25% of actual effort.

Appendix D contains these statistics for each software class analyzed.

The following applies to the Wilcoxon Signed-Rank Test (Mendenhall, 1990: 680-681):

1. H_o: the population distributions for actual effort and estimated effort are identical (the null hypothesis).

2. H_a: the two distributions differ in location (the alternative hypothesis).

3. The test statistic is: $$T = \min(T^-, T^+) \tag{10}$$

where T^- = sum of ranks of negative differences (actual effort - estimated effort) and T^+ = sum of ranks of positive differences.

4. If $T \leq T_o$ then we reject H_o and say that the distributions are not identical in location.

5. Comparing the values of T^- and T^+ to T_o allows to draw conclusions concerning the relative location of the two distributions. In other words, this "one-tailed test" allows us to detect the direction (positive or negative) of bias, if it exists.

Appendix E contains the results of the Wilcoxon test for each software class.

I also expect to utilize the Statistical Analysis Software (SAS) program available on the AFIT mainframe system to compare regression results with those of the DBMS for at least one software class. This will be done in an effort to determine the validity of relying upon the RMS statistic.

Figure 1 is a diagram which summarizes the 13 steps used to calibrate SASET. Each steps is discussed in more detail below the figure. It must be noted that many of the manipulations are not specified in any of the documentation and were acquired by the author via a telephone conversation with one of the model developers, Mr. Rick Maness. Also to be noted is the value of using *in-house* historical data to estimate the cost of a future *in-house* development project.

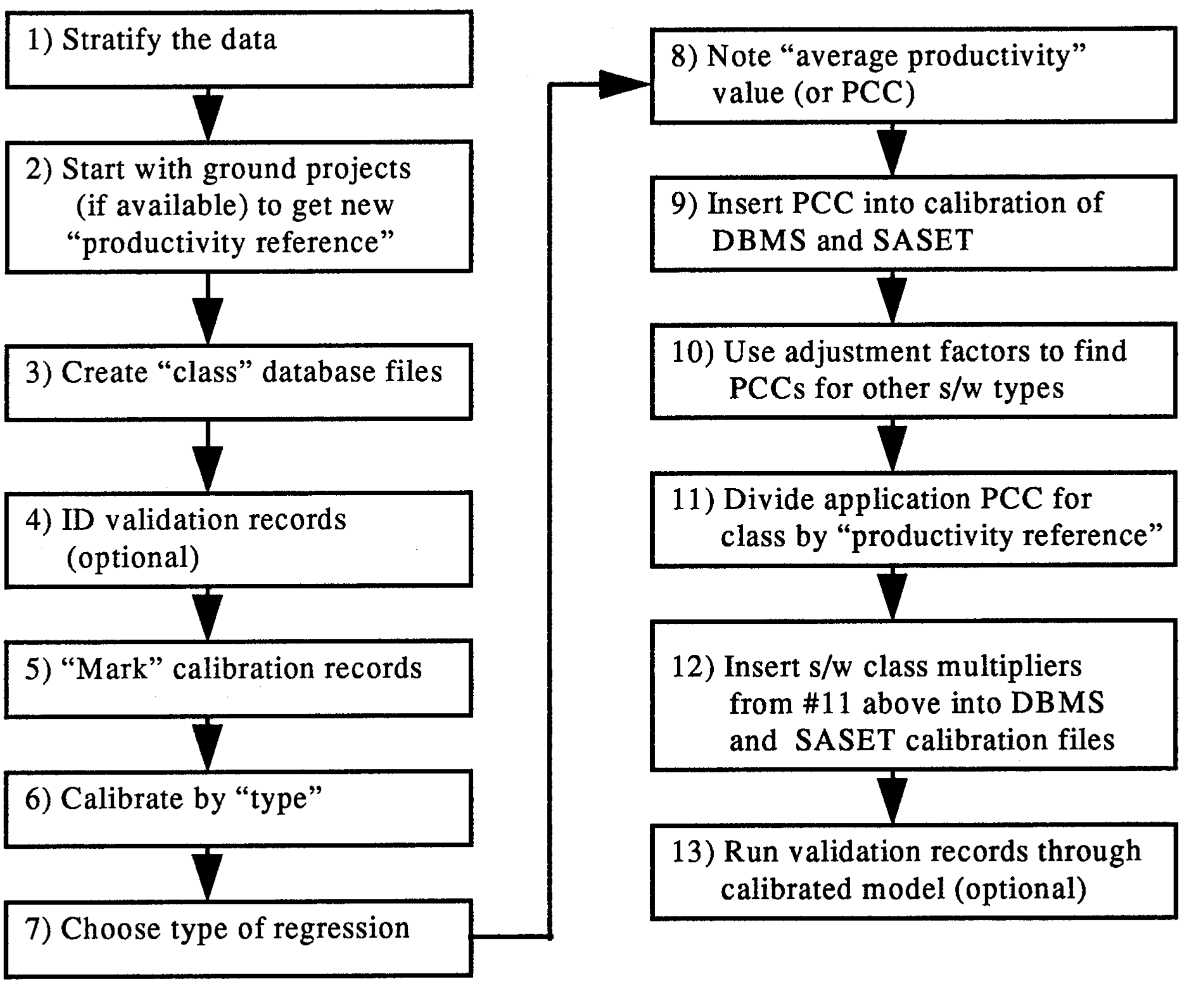

FIGURE 1. SASET Calibration Steps

The specific steps are as follows:

1. Stratify the data into groups you wish to analyze; the major divisions are necessarily driven by SASET's structure (i.e. class & type). Within these you can further stratify the data by any other characteristics you want, such as language, development standard, etc. (this assumes that you have enough records to do so)

2. If you have "ground" projects you should run these first in order to get the "productivity reference" value

3. Enter the data for all the records of a given "class" into a database file in the DBMS

4. Determine which records will be used for calibration (the rest will be used for validation); run the validation records through uncalibrated SASET, if desired, in order to make comparison and determine the usefulness of the calibration

5. "Mark for calibration" the records identified in #2 above; if no validation is desired then just select all the records

6. Run the automatic regression by choosing "calibrate" by "type"

7. Determine which type of regression you want to use (see discussion in chapter 4)

8. Write down "average productivity" value shown when the desired type of regression is highlighted (do not choose more than one regression type at a time)

9. The value from #8 above will be used in the PCC table of the calibration files of the DBMS and SASET; for the DBMS choose "Calibrate" then "Page 1" and for SASET choose "Calibrate" then "Software Development" the "PCC/LOC SCH/SM Equivalent" to input the PCCs

10. Depending on the "type" of software for which a value has been found one must make adjustments to the other types; for example (the most typical case in this research) is that you have mostly application records and thus get a productivity number for the application software of a given class. You can make adjustments to this value to come up with multipliers for "system" and "support" type software if you do not have enough records representing these types (see chapter 4)

11. Divide the application productivity value for the class of software you are analyzing by the "productivity reference" value in the DBMS calibration file, page one; this should be the "ground - application" productivity value

12. The value from #11 above will be the new s/w class multiplier; enter this value into the DBMS and the SASET calibration files; for the DBMS choose "Calibrate" then "Page 1" and for SASET choose "Calibrate" then "Software Development" then "Tier and Life Cycle Factors" to change the S/W Class Multipliers (or factors)

13. Run the validation records through the calibrated SASET and, if desired, compare the results to the estimates made through the uncalibrated SASET.

NOTE: If it is not desired to make a pre-calibration to post-calibration comparison (or the data set is too small to allow for this), eliminate the validation steps #4 and #13 listed above. All other steps should still be followed.

Results

The result of this research will provide SMC with a cost estimating model that is tailored/calibrated to their specific operating environment (represented by the historical

data in the SMC SWDB). The only inputs which will be varied in SASET to improve its accuracy are the PCC values and Software Class Effort Multipliers. However, if it is determined that the calibrated model is not more accurate than the uncalibrated model, there will be no justification for using the calibrated model. Again, note that improvement in accuracy does not guarantee that the calibrated model will accurately predict future costs for systems that are unlike the systems used for calibration. In fact, a calibrated model should be *assumed* to have some error in its estimates and should be evaluated for risk.

The primary goal of this research is simply to provide an improved model to the DoD for estimating effort for future software developments; secondary goals are to provide DoD with PCCs for the other classes of software represented in the database and to provide the reader with a relatively simple and clear reference on how to calibrate SASET using the DBMS.

IV. FINDINGS

Overview

This chapter will present the results of the research effort. Specifically, it will discuss assumptions made about missing data, adjustments made to the data, model-specific peculiarities encountered, and the comparison of uncalibrated model estimates against calibrated model estimates.

The Data

Identification. The goal was to calibrate homogenous groups of projects in order to come up with meaningful and representative PCCs. This was done by determining which characteristics were important and seeking out records with these characteristics in common.

The SWDB Report Writer allows the user to select from a large number of information fields to report (although there is a maximum of nine fields, in landscape mode, per report). The fields generated for this research were: record number (as the "keyfield"), application, program language, development country, development standard, level of complexity, normalized effective size, and normalized effort. One other field ("development phases included") was extracted manually since it was not listed as an available field within the Report Writer. Assumptions made for each of these fields in the absence of information are discussed in the next section.

The first step was necessarily to query the database by the primary stratification characteristic which was the "class" of software. Only after obtaining records which were from the class being analyzed could the Report Writer function be used to extract the fields mentioned above. Within each class, the records were substratified by "type" (system, application, and support). Determination of project type is discussed below.

The "record number" was designated as the keyfield simply for identification purposes. This field was the only field which was unique for each record.

The "application" field was used to determine the type of software represented by the record, with the help of Mr. Tom Phigetti, the author's point of contact at Martin-Lockheed. As it turned out the majority of the records for all the classes were of the "application" type. It should be noted that the term "application" is used in a different sense in the SWDB than in SASET.

The effect of "program language" was also analyzed. No distinction could be made among HOL in the sizing input windows. However, SASET and the DBMS do distinguish between high order language (HOL) and assembly. Distinguishing assembly programs from HOL programs was found to have an impact on estimates and regression results.

The "development country" was only a distinguishing factor for the "unmanned space" class. It was decided that only projects developed within the US would be used; some European records are contained in the database. This was done in an attempt to achieve as much homogeneity possible; the European records contained minimal

information and possibly used different development standards, labor hour standards and development techniques.

The reason "development standard" was deemed important is that it affects the way total effort is broken out among the development phases. This is accounted for in the percentage breakout of effort under the "budget information" window of the DBMS as well as in Tier 1 and Tier 3 of SASET. This breakout is also affected by the phases reported and is the reason for identifying the ninth field as "development phases included" (see Appendix B).

TABLE 2. Development Standard Effort Percentage Breakouts

Standard	Requirements	Design	Code	Test
2167A w/o requirement	25% 0%	24% 32%	20% 27%	31% 41%
483/490 w/o requirement	23% 0%	22% 29%	20% 26%	35% 45%
MCR/"other"	4.7%	26%	55.7%	13.7%

"Level of complexity" was originally considered an important information field but it was eventually decided that it should not be considered for the reasons discussed in the next section, "Assumptions."

"Normalized effective size" and "normalized effort" were obviously crucial pieces of information to the analysis and are discussed more fully in the section below entitled "Normalization." Appendix A shows the fields discussed above for each of the records used in the analysis.

Table 3 below shows the number of calibration points and validation points acquired for each class.

TABLE 3. Number of Data Points

	Space	Missile	Avionics	Commercial	Mobile	Ground
Calibration	26	4	8	3	10	49
Validation	13	0	1	0	4	24
Total	39	4	9	3	14	73

Recall that a minimum of five records are required for calibration by the DBMS. Therefore, the missile and commercial classes could not be calibrated. Refer to the "Research Design" section of Chapter 3 for the method used to divide data points between calibration and validation.

Assumptions. Some assumptions were necessarily made in the absence of data. Only records with SLOC and effort information were used. However, many of the records were missing other information needed for the Tier 1 and Tier 3 inputs. It was decided that it would be better to assume all inputs in these two tiers were nominal (default values); other than class and development standard in Tier 1 and documentation level in Tier 3 (which is driven by the development standard used).

Besides the fact that the majority of the records were missing information in most fields, this assumption of default input values was also made due to the inherent subjectivity in specifying some of the inputs such as level of complexity, programmer capability, etc. Hopefully, ignoring these inputs will prevent introducing bias and therefore result in a more general (and useful) calibration parameter. The other reason for

this assumption is the fact that SASET is not very sensitive to inputs other than SLOC. In fact, SASET's core estimate is simply (Coggins, 1993: 58):

$$\text{Effort (in manhours)} = \text{SLOC} * \text{S/W Class Multiplier} * \text{PCC for S/W Type} \quad (11)$$

This lack of sensitivity to other inputs is evident in the DBMS as well, which only allows for a general "complexity" input in addition to SLOC and effort. In fact, this input is used as a field upon which to calibrate but does not seem to affect any computations. Yet it should be stated that when estimating a new project all known inputs should be adjusted in order to fully utilize the model's capabilities.

Assumptions made for each of the fields discussed in the previous section, "Identification," were as follows:

- obviously no assumption could be made for the "keyfield" of record number
- if the "application" field descriptor could apply to more than one type of software, it was assumed to be Application type (several records which were definitely System or Support were not used since there were not enough records in any of the classes for a separate calibration of either of these two types)
- if the "programming language" was not reported it was assumed to be HOL
- "development country" was reported for all records
- "development standard" was assumed to be DoD-Std-2167A if none was reported and was assumed to follow the breakout specified by Mrs. Sherry Stukes of MCR, Inc. (bottom row in Table 2) if it was reported as "other"
- as discussed above, "level of complexity" was dismissed from consideration

- no assumptions could be made for "normalized effective size" or "normalized effort"

- if the "development phases included" field was blank, it was assumed that all phases applied. For the majority of the records that did contain this information, they either included requirements through testing or design through testing. Refer to the table above to observe the differences in percentages for these two scenarios.

Since the data was gathered at the CSCI level, only one CSCI per record was used in SASET to compute estimates. The Coggins and Russell thesis discusses the impact of multiple CSCIs on effort estimates in SASET (Coggins, 1993: 73). Coggins and Russell also mention that SASET's CSCI SLOC range is from about 500 to 120,000 (Coggins, 1993: 75). However, that is due to the database to which SASET is originally calibrated. The author believes that the range of the database used to calibrate SASET determines the range for which the model is valid.

Normalization. The SWDB Report Writer function provided the capability to determine the "normalized effective size" and the "normalized effort." The size was automatically normalized by the SWDB in order to account for differences between new, modified, and reused code. This was done by computing the "equivalent new SLOC" which was set equal to 40% of re-design, 25% of re-code, and 35% of re-test software. If all of the SLOC used was new, then the equivalent new SLOC equaled the new SLOC. The normalized value (equivalent SLOC) was input as "new code" in both SASET and

DBMS. If it had not been entered as "new code," SASET would make automatic adjustments on its own and thus distort the "equivalent" size data.

The effort was normalized in order to account for differences in the number of man-hours per manmonth; they were converted to the 152 manhours per manmonth standard. Normalized effort was used for all calculations.

The Model

DBMS. The important items in the DBMS were specifying whether the program was written in HOL or assembly and breaking out the effort by phase based upon the development standard used (see Table 2). The effort was not broken out by "organization" since the majority of the records were missing this information. It is uncertain what kind of impact this would have but based on discussion with Martin-Lockheed and MCR personnel it was deemed much more important to break effort out by phase.

Analysis was performed using the Statistical Analysis Software (SAS) program to determine the ability to rely on the RMS statistic (see Appendix F). This analysis was performed for the ground-application records and only considered linear, power, and logarithmic regressions. Several of the statistics reported in the Analysis of Variance (ANOVA) table created by SAS indicated that the linear regression was the best fit line. The RMS value for the linear regression type was lower than the other types analyzed. Therefore, it was concluded that, in addition to the graphs, the user probably can safely

rely on the RMS value to decide which type of regression to accept (with the caveats mentioned in Chapter 3).

Due to the fact that SASET uses ground software as a reference point in estimating effort, it was necessary to perform regression of this class of software first. Specifically, ground "application" software was analyzed in order to find the "productivity reference" needed for all other calculations (as per Mr. Maness). Multipliers used to determine the PCCs for the other types of software were found by applying the following rule of thumb, also specified by Mr. Maness:

$$\text{SYSTEM} = 1.2 \times \text{APPLICATION} \tag{12}$$

$$\text{SUPPORT} = .87 \times \text{APPLICATION} \tag{13}$$

These heuristic multipliers attempt to adjust for the relative effort required by each type of software of a given class, using ground-application as the reference point. The calibration constants are listed below in Tables 4 and 5 below.

<u>SASET</u>. The only inputs varied within SASET when calculating pre-calibration effort for the validation subset were: class of software and development standard in Tier 1, and the "software documentation" element in Tier 3 (which is driven by the development standard used). Of course, the PCC values were also changed in the calibration file once calibration was performed in order to compute the post-calibration estimates of effort.

One observation made by the author is that, while DBMS allows for various types of regression equations to be employed, SASET estimates are based on linear regression.

Thus, if the true line for the data is found to be curvilinear when using the DBMS, SASET can not model that curvilinear pattern (all inputs other than size held constant).

The Results

PCCs & Software Class Multipliers. The default PCCs and multipliers for each class and type of software are shown in Table 4. The PCCs and multipliers for each of the software classes and types resulting from calibration are shown in Table 5.

The general method used in calculating the Software Class Effort Multiplier for a given class and type of software (as per Mr. Maness) is:

$$\text{S/W Class Mult.} = \frac{\text{Average Productivity for Class (or PCC)}}{\text{Productivity Reference}} \quad (14)$$

TABLE 4. Default PCCs

	Ground	Space	Avionics	Mobile	Commercial	Missile
System	3.3	3.3	3.3	3.3	3.3	N/A
Application	1.9	1.9	1.9	1.9	1.9	N/A
Support	.85 1.1*	.85 1.1*	.85 1.1*	.85 1.1*	.85 1.1*	N/A
S/W Class Multiplier	1.0	2.3	1.8	1.35	.75	N/A

NOTE: The Support PCC is .85 in SASET but is 1.1 in DBMS; the "mobile" class was considered to be equivalent to the Ship/Submarine class in SASET; the "missile" class does not exist in SASET; the "space" class refers to *unmanned* space flight (not *manned* space flight); the "productivity reference" value is 2.2.

We see from Table 5 that there is a substantial difference between the default PCCs and the calibrated PCCs. Notice that the PCCs are all the same in the default table;

TABLE 5. Calibration PCCs

	Ground	Space	Avionics	Mobile	Commercial	Missile
System	1.3724	0.2570 (4.7705)	2.0804	2.2390	N/A	N/A
Application	**1.1437**	0.2142 (3.9754)	1.7337	1.8658	N/A	N/A
Support	0.9950	0.1864 (3.4586)	1.5083	1.6232	N/A	N/A
S/W Class Multiplier	1.0000	0.1873 (3.4759)	1.5159	1.6314	N/A	N/A
Correlation	0.931	0.324	0.777	0.643	N/A	N/A

NOTE: The "productivity reference" value is shown in bold. Also, the PCCs in parentheses were arrived at using logarithmic regression in the DBMS; all others were arrived at using linear regression.

the only difference is found in the S/W Class Multiplier values. Since the "missile" class is not represented in SASET, there are no default values for this class.

The calibrated PCC table shows the new productivity reference value to be 1.1437 hours/ SLOC. It seems reasonable to observe improved productivity reflected in the more current records of the SWDB as opposed to those in SASET's internal database. The assumption for this argument is that new products and processes used in software engineering reduce labor intensity in addition to improving the end product.

As can be seen from Tables 4 and 5, improved productivity is observed for all the classes except "mobile." Changes in the PCC and in the productivity reference result in changes in the S/W Class Multipliers. The only multiplier which increased is for the "mobile" class. Of course, the multiplier should always be 1.0 for the "ground" class. The "correlation" values shown were computed by the DBMS and reflect the overall relation between SLOC and effort for a given class.

Also note the values for the "space" class in the calibrated PCC table. The top values represent results using linear regression while the values in parentheses represent results using logarithmic regression. The low correlation value of 0.324 made it difficult to determine a meaningful regression line. It was noticed that the logarithmic regression had the lowest RMS value but the linear regression RMS was not much higher. Also, both graphs appeared to be equally valid. Therefore, both regression types were analyzed. Finally, note that there were not enough useable "commercial" and "missile" records in the database to perform calibration. All the statistical results for each class analyzed are discussed in the sections that follow.

Pre-calibration. SASET's pre-calibration accuracy for each class of software is given below (also refer to Appendix D). "Application" projects were the overwhelming majority and their PCCs were therefore calculated directly. PCCs for the other types of software were calculated using the heuristic multipliers specified by Mr. Maness (refer to equations 12 and 13).

Chapter 3 describes the equations used to calculate the statistics contained in Table 6. As stated in the footnote of Table 6, the mean, variance, and standard deviation values refer to the *difference* between actual effort and estimated effort. This helps identify bias in the model in terms of consistent overestimation or underestimation.

The negative values for the "mean" statistic for all the classes indicate that, on average, actual effort is lower than estimated effort for each of the classes. In other

TABLE 6. Pre-Calibration Statistics

	Ground	Space (linear)	Avionics	Mobile
Mean	-374	-1745	-430	-99
Var	235,349	7,744,016	N/A	78,679
Std Dev	485	2782	N/A	280
MMRE	10.04	5.54	1.76	5.61
Wilcoxon	+ Bias	+ Bias	N/A	N/A
% Descrip.	0%	23%	0%	25%

NOTE: Mean, Var, and Std Dev are for the difference between actual effort and estimated effort.

words, SASET appears to have a tendency to overestimate. This conclusion is somewhat affirmed by the results of the Wilcoxon test for the ground and space categories. Unfortunately, the Wilcoxon test could not be used for any of the other software classes due to their small sample sizes.

The variance and standard deviation statistics refer to the spread of the distributions of the differences and are more meaningful when compared with the post-calibration results shown in Table 7.

Since MMRE stands for Mean Magnitude of Relative Error, it is clear that the lower this value is, the better the performance of the model. Again, these values should be compared with those in Table 7.

The last measure, the "% Description" statistic (which is the percentage of estimates that are within 25% of the actual), is commonly used in the software industry and has thus become a standard one to report. Although it may be of questionable academic use, it does give a rough idea of model performance. However, it should be

emphasized that no one statistic provides the total "picture;" that is why they are all being reported and analyzed simultaneously.

It is important to mention that because of the fact that there was only one "avionics" validation point, several of the statistics could not be computed or became meaningless.

Table 6 gives us an impression of SASET's performance prior to calibration. Yet, alone it is of limited value. Its information must be compared to results achieved after calibration in order to be of use for this research. These "post-calibration" results are reported below.

Post-calibration. Statistics for the calibrated model are listed in Table 7 below. There are several important points to be made about Table 7. Note that results for both regression types for the "space" class have been reported. The results for this class will be discussed last. All the statistics which were calculated before calibration were also calculated after calibration.

TABLE 7. Post-Calibration Statistics

	Ground	Space (linear)	Space (log)	Avionics	Mobile
Mean	-37	333	-6281	-55	8
Var	94,006	180,729	88,880,541	N/A	72,354
Std Dev	307	425	9428	N/A	269
MMRE	5.82	.94	19.54	.22	3.57
Wilcoxon	No Bias	- Bias	+ Bias	N/A	N/A
% Descrip.	38%	0%	0%	100%	0%

The statistics for the "ground" class all indicate that calibration has markedly improved the model's ability to estimate for this class. We see that the mean difference is now much closer to zero and is one-tenth of its pre-calibration value. This indicates a significant reduction in bias; which is also indicated by the Wilcoxon test.

The variance is also greatly reduced after calibration is performed. This indicates that the distribution of the differences between actual and estimated effort is narrower or "tighter" than before calibration.

Finally, the "% Descrip" increased from 0% to 38%. Although a greater percentage would be desirable, the important fact is that there was improvement after calibration. Recall that the purpose of this analysis is to investigate improvement in estimation capability rather than report the model's absolute accuracy.

These favorable statistics were a welcome result since SASET uses "ground" as the basis for estimating for all the other classes. The high correlation between SLOC and effort found among the records for this class likely played a significant role in achieving these superior results.

Although many of the statistics for the "avionics" class could not be computed, due to the fact that only one validation point was available, the ones which were computed indicate that calibration improved the model's performance.

Note that the mean decreased from -430 to -55; significantly closer to zero. Unfortunately the Wilcoxon test could not be performed to further analyze bias. The MMRE also decreased; it improved from 1.76 to .22 after calibration. The "% Descrip." was the only other statistic which could be computed. Although it appears as a very

impressive improvement, recall that only one record was used for validation; thus the 100% represents that single record coming within 25% of the actual.

Related to the fact that one record was used for validation and eight were used for calibration is the idea that one should be wary of results from such a small collection of data. However, if that is all that is available then one has no choice but to rely on the results until more data becomes available.

The "mobile" class also reflects improvement, although not as impressive as the other two classes discussed above. The results show moderate improvement in all the statistics. Again this is likely related to the correlation reported in Table 5. The "mobile" class had a moderate correlation of 0.643.

Finally, the "space" class was a bit more difficult to analyze. As mentioned previously, the low correlation of 0.324 made it difficult to determine which type of regression was most appropriate. Although both the linear and logarithmic regressions resulted in suspicious estimates (refer to Appendix D), the pre-calibration estimates were also suspect (in light of the actuals).

The final analysis is that the linear regression calibration resulted in the *least* suspicious estimates (relatively speaking). It also had the best statistics; with the exception of "% Descrip." The mean of 333 was much lower than the pre-calibration and the logarithmic calibration means. The variance was significantly reduced by the linear regression, while it was increased by the logarithmic regression. The MMRE of .94 was low in absolute terms as well as in relative terms. But, as mentioned, the % Description

decreased from the pre-calibration value of 23% to 0%. This emphasizes the importance of basing analysis upon more than one statistic.

Perhaps a reason for these unexciting results is that effort for writing software for this class is not strongly related to the SLOC or perhaps there are one or two anomalies among the records that are distorting the true relationship. However, with the information available, the author could not justify further eliminating any of the records used.

Comparison. Table 8 contains the same information reported in Tables 6 and 7 but simply allows for easier comparison of changes in the statistics after calibration. Refer to the sections above for discussions concerning the differences between the pre-calibration and post-calibration values.

Summary. The intent of this chapter was to present the quantitative results of the analysis and make detailed observations about these results. One important observation is that the strength of correlation between SLOC and effort for the database used for

TABLE 8. Summary of Statistics

		Ground	Space (linear)	Space (log)	Avionics	Mobile
Mean	Pre	-374	-1745	-1745	-430	-99
	Post	-37	333	-6281	-55	8
Var	Pre	235,349	7,744,016	7,744,016	N/A	78,679
	Post	94,006	180,729	88,880,541	N/A	72,354
Std Dev	Pre	485	2782	2782	N/A	280
	Post	307	425	9428	N/A	269
MMRE	Pre	10.04	5.54	5.54	1.76	5.61
	Post	5.82	.94	19.54	.22	3.57
Wilcoxon	Pre	+ Bias	+ Bias	+ Bias	N/A	N/A
	Post	No Bias	- Bias	+ Bias	N/A	N/A
% Descrip.	Pre	0%	23%	23%	0%	25%
	Post	38%	0%	0%	100%	0%

calibration has a major impact on the ability to use the DBMS to derive representative PCCs. Unfortunately, the "space" class, which is of most interest to the SMC, was the most difficult class to calibrate in this research effort. Its results are the most suspect. However, the "ground" class, which is the foundation for all estimates in SASET, had the most encouraging results.

The next chapter will convey general conclusions about the research and make recommendations for future research.

V. CONCLUSIONS AND RECOMMENDATIONS

Conclusion

The objective of this research effort was to provide the USAF SMC, and more generally the DoD, with general calibration parameters to be used in the SASET software cost estimating model. These calibration parameters were to be based on a large, current database (SWDB) maintained by SMC.

In general, the results were encouraging. Calibration of SASET appeared to significantly improve the model's accuracy, as represented by the statistics used in this research. Yet the results were not unanimous. As mentioned in Chapter 4, the results for the "space" class were suspicious. Although there seemed to be enough records available to perform useful analysis, the relationship between SLOC and effort in this software class was not strong enough to result in a clear improvement. Despite the improvement indicated by the statistics after using the linear regression PCC, a more subjective analysis of the values estimated by the model led to the conclusion that the results were suspect.

However, several of the other classes (military ground, avionics, and military mobile) seemed to be successfully calibrated and, in the absence of PCCs developed in-house, use of the reported PCCs should help improve SASET's performance in estimating new projects within those classes. Unfortunately, the missile and commercial software categories could not be analyzed due to a lack of useful data. This indicates the

importance of having records with sufficient detail as well as collecting as many records as possible. In other words, the quality of the data points is just as important as the quantity.

Another point is the importance of thorough documentation. SASET's documentation has a reputation of being sparse, and this author found this to be true for SASET and the DBMS; both the printed documentation and the on-screen documentation. It is unreasonable to expect the average user to perform the research performed here in order to calibrate the model. The documentation should support the product and not leave the user with fundamental questions which can only be answered by an engineer who helped develop the product. It is true that fully detailed documentation would be very difficult due to SASET's openness (or ability to change virtually all default values) but calibration capabilities become useless if one does not fully understand how to exploit them. This research will hopefully reach those SASET users who would like to know what equations and what specific steps are necessary to calibrate the model.

Recommendations

Although lack of data has been identified as a problem by many other researchers, this author would like to reemphasize the fact that there truly needs to be more dedicated collection of information in order for calibration efforts to be truly useful. Reporting accurate and detailed information on current projects needs to be emphasized and enforced. After all, historical information is the basis for regression-based calibration. Without useful and valid data, and enough of it, the usefulness and validity of the resulting calculations consequently become questionable.

Future research may want to analyze whether breaking out effort by "organization" as well as by phase has an impact on results within SASET. Also, it may be interesting to examine whether the heuristic multipliers used to adjust for the type of software, offered by the creators of the model, are still valid. Of course, both of these analyses would require the availability of records containing the necessary information. Other studies may look into: whether the method used by the SWDB to normalize SLOC is valid, how much of an impact changing some of the system attributes (Tier 3) would have on results, whether European software development is similar enough to combine their records with US data records, and whether the DBMS would prove to be more useful if SASET could model the regression equations contained within the tool.

Improvements in software development techniques and processes are likely having an effect on the way software is produced as well as on the traditional distinctions made between different types of projects. For example, these improvements may be decreasing the effort required to reuse code or diminishing the difference in productivity among the different types and classes of software. It may be that such changes have eliminated the applicability of some inputs, within SASET and other software cost estimating models, which have traditionally been identified as key cost drivers. A study of major changes in the software development industry and the effects of these changes on productivity and current estimating equations would be of great interest, especially in light of the increasing portion of system costs which software represents.

Appendix A: Data Records

Military Ground

Number of records included in search: 83

Keyfield	Application	Prog Lang	Dev Coun	Dev Std	Lv of Cmplx
00000001	Command/Control		USA		
00000002	Test	FORTRAN 1%	USA		Routine
00000004	OS/Executive		USA		
00000005	Simulation		USA		
00000006	Test		USA		
00000007	Command/Control	Assembly 35% FORTRAN 65%	USA		Simple
00000009	Command/Control	Assembly 8% FORTRAN 92%	USA		Routine
00000023	Test	Atlas 100%	USA		Difficult
00000024	Test	Atlas 100%	USA		Difficult
00000025	Simulation	Assembly FORTRAN 100%	USA		Difficult
00000026	Test	Atlas 100%	USA		Difficult
00000028	Mission Planning	FORTRAN 5% PLI 95%	USA		Routine
00000030	Test	Assembly 28% PASCAL 72%	USA		Complex
00000039	CAD	FORTRAN 100%	USA		Simple
00000040	CAD	FORTRAN 100%	USA		Simple
00000041	Simulation	FORTRAN 100%	USA		Routine
00000048	Mission Planning	FORTRAN 100%	USA		Simple
00000050	Command/Control	FORTRAN	USA		Complex
00000053	Test	Assembly 100%	USA	MIL-STD-483/490	Routine
00000054	Signal Processing	Assembly 7% PASCAL 93%	USA	MIL-STD-483/490	Routine
00000055	Test	Assembly 100%	USA	MIL-STD-483/490	Difficult
00000056	Database	Assembly 100%	USA	MIL-STD-483/490	Routine
00000058	Test	Atlas 100%	USA	MIL-STD-483/490	Routine
00000059	Test	Atlas 100%	USA	MIL-STD-483/490	Routine
00000061	MMI/Graphics	Assembly 93% FORTRAN 7%	USA	MIL-STD-483/490	Routine
00000062	Utilities	Assembly 7% FORTRAN 93%	USA	MIL-STD-483/490	Simple
00000063	Test	FORTRAN 100%	USA	MIL-STD-483/490	Simple
00000068	Training	Assembly 35% FORTRAN 65%	USA	DoD-STD-1679	Routine
00000072	Process Control	Assembly 30% PASCAL 70%	USA		Difficult
00000073	OS/Executive	Assembly 19% PASCAL 81%	USA		Difficult

Military Ground

Number of records included in search: 83

Keyfield	Application	Prog Lang	Dev Coun	Dev Std	Lv of Cmplx
00000120	Command/Control		USA		
00000121	Utilities		USA		
00000122	Utilities		USA		
00000123	Database		USA		
00000124	Command/Control		USA		
00000125	Utilities		USA		
00000126	Signal Processing		USA		
00000127	Signal Processing		USA		
00000128	Utilities		USA		
00000129	Test		USA		
00000130	Signal Processing		USA		
00000131	Signal Processing		USA		
00000132	Signal Processing		USA		
00000133	Signal Processing		USA		
00000134	Signal Processing		USA		
00000135	Signal Processing		USA		
00000136	Signal Processing		USA		
00000137	Signal Processing		USA		
00000138	Signal Processing		USA		
00000139	Test		USA		
00000140	Signal Processing		USA		
00000141	Command/Control		USA		
00000142	Signal Processing		USA		
00000143	Signal Processing		USA		
00000144	Signal Processing		USA		
00000145	Command/Control		USA		
00000147	Signal Processing		USA		
00000148	Test		USA		
00000150	Command/Control		USA		
00000151	Database		USA		
00000152	Command/Control		USA		
00000153	Signal Processing		USA		
00000154	Signal Processing		USA		
00000155	Command/Control		USA		
00000301	Training	Assembly 5% FORTRAN 95%	USA	MIL-STD-483/490	Simple
00002497	Diagnostics	Ada 100%	USA	Other	Difficult
00002498	MIS	Ada 90% Other 10%	USA	Other	Routine

Military Ground

Number of records included in search: 83

Keyfield	Application	Prog Lang	Dev Coun	Dev Std	Lv of Cmplx
00002501	Command/Control	Ada 98% Assembly 2%	USA	Other	Routine
00002510	Command/Control	C 100%	USA	DoD-STD-1703	Routine
00002517	Command/Control	Assembly 50% C 25% FORTRAN 25%	USA	Other	Routine
00002519	MIS	Assembly 1% COBOL 65% Other 34%	USA	DoD-STD-2167A (Full)	Simple
00002520	MIS	Assembly 1% COBOL 61% Other 38%	USA	DoD-STD-2167A (Full)	Simple
00002521	MIS	COBOL 82% Other 18%	USA	DoD-STD-2167A (Full)	Simple
00002522	MIS	COBOL 34% Other 66%	USA	DoD-STD-2167A (Full)	Simple
00002523	MIS	Basic 5% C 10% COBOL 46% Other 39%	USA	DoD-STD-2167A (Full)	Simple
00002524	MIS	C 2% COBOL 47% Other 51%	USA	DoD-STD-2167A (Full)	Simple
00002525	MIS	COBOL 50% Other 50%	USA	DoD-STD-2167A (Full)	None
00002526	MIS	COBOL 100%	USA	DoD-STD-2167 (Tailored)	Complex
00002527	MIS	COBOL 100%	USA	DoD-STD-2167 (Tailored)	Complex
00002528	MIS	COBOL 100%	USA	DoD-STD-2167 (Tailored)	Complex
00002610	MIS	COBOL 100%	USA	DoD-STD-2167 (Tailored)	Complex
00002611	MIS	COBOL 100%	USA	DoD-STD-2167 (Tailored)	Complex
00002612	MIS	COBOL 100%	USA	DoD-STD-2167 (Tailored)	Complex

Military Ground (cont'd)

Number of records included in search: 83

Keyfield	Norm Eff Sz	Norm Effrt
00000001	1500	9
00000002	28805	232
00000004	46000	84
00000005	16000	49
00000006	41000	66
00000007	45057	120
00000009	128200	517
00000023	5200	40
00000024	18000	226
00000025	111995	720
00000026	4170	21
00000028	112917	841
00000030	41000	152
00000039	17000	19
00000040	23483	54
00000041	9500	22
00000048	136227	326
00000050	144000	684
00000053	21122	48
00000054	45035	127
00000055	260882	6363
00000056	22150	22
00000058	28191	606
00000059	15554	139
00000061	72716	107
00000062	16428	36
00000063	11753	22
00000068	37457	161
00000072	25160	45
00000073	31661	60
00000120	25842	95
00000121	66548	66
00000122	31673	46
00000123	47600	176
00000124	23881	139
00000125	20042	78
00000126	47965	165
00000127	16016	13

Military Ground (cont'd)

Number of records included in search: 83

Keyfield	Norm Eff Sz	Norm Effrt
00000128	40141	97
00000129	22516	44
00000130	71851	738
00000131	29147	192
00000132	46595	278
00000133	123710	645
00000134	44527	228
00000135	23787	264
00000136	12121	154
00000137	60233	274
00000138	14389	190
00000139	38634	237
00000140	70020	6
00000141	162039	322
00000142	28782	348
00000143	23703	86
00000144	29802	145
00000145	18560	101
00000147	31720	192
00000148	300000	2551
00000150	21681	100
00000151	38174	190
00000152	69772	286
00000153	11534	149
00000154	8965	109
00000155	8398	74
00000301	76566	408
00002497	10000	76
00002498	100000	16
00002501	110400	356
00002510	43437	172
00002517	85382	167
00002519	419619	1560
00002520	419619	3370
00002521	97087	626
00002522	461426	2191
00002523	231018	949
00002524	363371	1298

Military Ground (cont'd)

Number of records included in search: 83

Keyfield	Norm Eff Sz	Norm Effrt
00002525	200000	993
00002526	6681	211
00002527	7457	235
00002528	21588	681
00002610	14536	458
00002611	11840	374
00002612	9899	312

Unmanned Space

Number of records included in search: 114

Keyfield	Application	Prog Lang	Dev Coun	Dev Std	Lv of Cmplx
00000003	Command/Control		USA		
00000029	OS/Executive	Ada 95% Assembly 5%	USA		Routine
00000038	Command/Control	Assembly 100%	USA		Complex
00000074	Command/Control	JOVIAL 100%	USA	MIL-STD-483/490	Difficult
00000075	Command/Control	JOVIAL 100%	USA	MIL-STD-483/490	Complex
00000076	Command/Control	JOVIAL 100%	USA	MIL-STD-483/490	Routine
00000077	Command/Control	JOVIAL 100%	USA	MIL-STD-483/490	Complex
00000078	Command/Control	JOVIAL 100%	USA	MIL-STD-483/490	Routine
00000079	Command/Control	JOVIAL 100%	USA	MIL-STD-483/490	Routine
00000080	Command/Control	FORTRAN 45% JOVIAL 55%	USA	MIL-STD-483/490	Routine
00000081	Command/Control	JOVIAL 100%	USA	MIL-STD-483/490	Routine
00000082	Command/Control	JOVIAL 100%	USA	MIL-STD-483/490	Routine
00000083	Command/Control	JOVIAL 100%	USA	MIL-STD-483/490	Simple
00000084	Command/Control		USA		
00000085	Command/Control		USA		
00000086	Signal Processing		USA		
00000088	Database		USA		
00000089	Mission Planning		USA		
00000090	Signal Processing		USA		
00000091	Signal Processing		USA		
00000092	Mission Planning		USA		
00000093	Command/Control		USA		
00000095	Command/Control		USA		
00000096	Signal Processing		USA		
00000097	Database		USA		
00000098	Mission Planning		USA		
00000099	Signal Processing		USA		
00000103	Command/Control		USA		
00000104	Signal Processing		USA		
00000105	Database		USA		
00000106	Mission Planning		USA		
00000107	Signal Processing		USA		
00000112	Command/Control		USA		
00000113	Command/Control		USA		
00000114	Database		USA		
00000115	Mission Planning		USA		
00000116	Mission Planning		USA		

Unmanned Space

Number of records included in search: 114

Keyfield	Application	Prog Lang	Dev Coun	Dev Std	Lv of Cmplx
00000117	Signal Processing		USA		
00000118	Signal Processing		USA		
00000119	Command/Control		USA		
00000305	OS/Executive	Ada 30% C 70%	USA	Commercial	Routine
00000306	OS/Executive	C 100%	USA	Commercial	Simple
00002516	Utilities	C 90% Machine 10%	USA	Other	Difficult
00002518	Command/Control	Assembly 16% C 83%	USA	Other	Difficult
00002529	Command/Control	Ada 100%	EUROPE	Other	None
00002531	Command/Control	FORTRAN 100%	EUROPE	Other	
00002532	Command/Control	FORTRAN 100%	EUROPE	Other	
00002533	Command/Control	Assembly 100%	EUROPE	Other	None
00002534	Command/Control	Ada 100%	EUROPE	Other	
00002536	Command/Control	Other 100%	EUROPE	Other	
00002539	Command/Control	Assembly 100%	EUROPE	Other	
00002540	Command/Control	Assembly 100%	EUROPE	Other	
00002542	Command/Control	Ada 100%	EUROPE	Other	
00002543	Command/Control	Other 100%	EUROPE	Other	
00002544	Command/Control	Ada 100%	EUROPE	Other	
00002545	Command/Control	FORTRAN 100%	EUROPE	Other	
00002546	Command/Control	Ada 100%	EUROPE	Other	
00002547	Command/Control	PASCAL 100%	EUROPE	Other	
00002548	Command/Control	Ada 100%	EUROPE	Other	
00002549	Command/Control	Other 100%	EUROPE	Other	
00002550	Mission Planning	PASCAL 100%	EUROPE	Other	
00002551	Mission Planning	FORTRAN 100%	EUROPE	Other	
00002552	Mission Planning	FORTRAN 100%	EUROPE	Other	
00002553	Mission Planning	PASCAL 100%	EUROPE	Other	
00002554	Mission Planning	Ada 100%	EUROPE	Other	
00002555	Mission Planning	Ada 100%	EUROPE	Other	
00002556	Mission Planning	Ada 100%	EUROPE	Other	
00002557	Mission Planning	Ada 100%	EUROPE	Other	
00002558	Message Switching	Ada 100%	EUROPE	Other	
00002559	Message Switching	Ada 100%	EUROPE	Other	
00002560	Message Switching	Other 100%	EUROPE	Other	
00002561	Message Switching	Other 100%	EUROPE	Other	
00002562	Message Switching	Assembly 100%	EUROPE	Other	
00002563	Message Switching	Assembly 100%	EUROPE	Other	

Unmanned Space

Number of records included in search: 114

Keyfield	Application	Prog Lang	Dev Coun	Dev Std	Lv of Cmplx
00002564	Message Switching	Assembly 100%	EUROPE	Other	
00002566	Signal Processing	Other 100%	EUROPE	Other	
00002567	Signal Processing	Assembly 100%	EUROPE	Other	
00002570	Signal Processing	C 100%	EUROPE	Other	
00002571	Signal Processing	Ada 100%	EUROPE	Other	
00002572	Signal Processing	Ada 100%	EUROPE	Other	
00002573	Signal Processing	Ada 100%	EUROPE	Other	
00002574	Signal Processing	C 100%	EUROPE	Other	
00002575	Signal Processing	C 100%	EUROPE	Other	
00002576	Signal Processing	Assembly 100%	EUROPE	Other	
00002577	Signal Processing	Assembly 50% FORTRAN 50%	EUROPE	Other	
00002578	Signal Processing	Assembly 50% C 50%	EUROPE	Other	
00002579	Signal Processing	Assembly 100%	EUROPE	Other	
00002580	Signal Processing	C 100%	EUROPE	Other	
00002581	Signal Processing	Assembly 100%	EUROPE	Other	
00002582	Signal Processing	Assembly 100%	EUROPE	Other	
00002583	Signal Processing	Other 100%	EUROPE	Other	
00002584	Signal Processing	Ada 100%	EUROPE	Other	
00002585	Signal Processing	C 50% PASCAL 50%	EUROPE	Other	
00002586	Signal Processing	PASCAL 100%	EUROPE	Other	
00002587	Signal Processing	Other 100%	EUROPE	Other	
00002588	Signal Processing	PASCAL 100%	EUROPE	Other	
00002589	Signal Processing	Other 100%	EUROPE	Other	
00002590	Signal Processing	FORTRAN 100%	EUROPE	Other	
00002591	Signal Processing	FORTRAN 100%	EUROPE	Other	
00002592	Signal Processing	Ada 100%	EUROPE	Other	
00002594	Signal Processing	FORTRAN 100%	EUROPE	Other	
00002595	Signal Processing	Other 100%	EUROPE	Other	
00002596	Signal Processing	Other 100%	EUROPE	Other	
00002597	Simulation	C 100%	EUROPE	Other	
00002598	Simulation	C 100%	EUROPE	Other	
00002599	Simulation	Other 100%	EUROPE	Other	
00002600	Simulation	Ada 100%	EUROPE	Other	
00002601	Simulation	Ada 100%	EUROPE	Other	
00002602	Simulation	FORTRAN 100%	EUROPE	Other	
00002603	S/W Development	C 100%	EUROPE	Other	

Unmanned Space

Number of records included in search: 114

Keyfield	Application	Prog Lang	Dev Coun	Dev Std	Lv of Cmplx
	Tools				
00002605	S/W Development Tools	Ada 100%	EUROPE	Other	
00002607	Other	COBOL 100%	EUROPE	Other	
00002608	Other	Ada 100%	EUROPE	Other	
00002609	Other	Ada 100%	EUROPE	Other	

Unmanned Space (cont'd)

Number of records included in search: 114

Keyfield	Norm Eff Sz	Norm Effrt
00000003	80000	583
00000029	2000	49
00000038	4671	61
00000074	11700	80
00000075	116800	912
00000076	14000	115
00000077	56200	523
00000078	48300	478
00000079	50300	432
00000080	69450	296
00000081	22900	164
00000082	16300	140
00000083	6800	57
00000084	6000	798
00000085	1950	204
00000086	6000	200
00000088	117000	244
00000089	225000	602
00000090	95000	1055
00000091	52275	1169
00000092	2920	75
00000093	250000	401
00000095	600	53
00000096	600	106
00000097	80000	530
00000098	90300	86
00000099	8000	234
00000103	600	7
00000104	600	191
00000105	21000	5
00000106	16300	206
00000107	8000	160
00000112	8290	1511
00000113	19500	1248
00000114	162945	235
00000115	13000	109
00000116	399635	1468
00000117	66843	652

Unmanned Space (cont'd)

Number of records included in search: 114

Keyfield	Norm Eff Sz	Norm Effrt
00000118	358000	765
00000119	278488	787
00000305	12810	143
00000306	9334	94
00002516	48941	167
00002518	17783	82
00002529	45000	90
00002531	130000	345
00002532	126000	244
00002533	16000	18
00002534	6000	9
00002536	22000	636
00002539	84000	793
00002540	18000	74
00002542	6000	56
00002543	11000	105
00002544	22000	118
00002545	19000	42
00002546	42000	85
00002547	100000	100
00002548	150000	222
00002549	21000	43
00002550	24000	89
00002551	19000	65
00002552	12000	30
00002553	35000	85
00002554	24000	31
00002555	83000	103
00002556	11000	12
00002557	11000	15
00002558	55000	292
00002559	2000	31
00002560	18000	145
00002561	47000	331
00002562	29000	234
00002563	17000	196
00002564	50000	278
00002566	5000	48

Unmanned Space (cont'd)

Number of records included in search: 114

Keyfield	Norm Eff Sz	Norm Effrt
00002567	13000	131
00002570	14000	66
00002571	3000	35
00002572	12000	28
00002573	4000	55
00002574	34000	181
00002575	9000	25
00002576	11000	202
00002577	22000	768
00002578	5000	63
00002579	32000	410
00002580	7000	93
00002581	30000	764
00002582	15000	313
00002583	62000	497
00002584	7000	12
00002585	14000	23
00002586	100000	186
00002587	32000	72
00002588	35000	128
00002589	10000	140
00002590	16000	59
00002591	10000	55
00002592	50000	113
00002594	45000	156
00002595	14000	58
00002596	40000	221
00002597	75000	130
00002598	14000	45
00002599	49000	526
00002600	3000	18
00002601	80000	197
00002602	50000	138
00002603	55000	225
00002605	12000	36
00002607	5000	37
00002608	55000	71
00002609	30000	60

Missile

Number of records included in search: 5

Keyfield	Application	Prog Lang	Dev Coun	Dev Std	Lv of Cmplx
00000015	Command/Control	Assembly 100%	USA		Routine
00000016	Command/Control	Assembly 100%	USA		Routine
00000017	OS/Executive	Assembly 100%	USA		Routine
00000027	Command/Control	JOVIAL 100%	USA		Complex
00000036	Command/Control	Assembly 100%	USA		Difficult

Missile (cont'd)

Number of records included in search: 5

Keyfield	Norm Eff Sz	Norm Effrt
00000015	8885	262
00000016	9025	91
00000017	1002	73
00000027	18933	1384
00000036	13658	480

Commercial

Number of records included in search: 3

Keyfield	Application	Prog Lang	Dev Coun	Dev Std	Lv of Cmplx
00000070	Training	Assembly 4% PASCAL 96%	USA		Routine
00000307	S/W Development Tools	C 100%	USA	Commercial	Difficult
00000309	Database	FORTRAN 80% Other 20%	USA		Simple

Commercial (cont'd)

Number of records included in search: 3

Keyfield	Norm Eff Sz	Norm Effrt
00000070	6386	43
00000307	21642	49
00000309	70000	222

Avionics

Number of records included in search: 9

Keyfield	Application	Prog Lang	Dev Coun	Dev Std	Lv of Cmplx
00000010	Command/Control	JOVIAL 95%	USA		
00000011	MMI/Graphics	JOVIAL 95%	USA		
00000012	MMI/Graphics	JOVIAL	USA		
00000013	Process Control	JOVIAL	USA		
00000014	Command/Control	JOVIAL 85%	USA		Complex
00000067	Signal Processing	JOVIAL 100%	USA		Simple
00000302	Diagnostics	Assembly JOVIAL	USA	DoD-STD-2167 (Tailored)	Complex
00000346	Command/Control	Other	USA		Complex
00002512	Command/Control	Ada 98% Assembly 2%	USA	DoD-STD-2167 (Tailored)	Difficult

Avionics (cont'd)

Number of records included in search: 9

Keyfield	Norm Eff Sz	Norm Effrt
00000010	43207	370
00000011	32878	198
00000012	22027	112
00000013	58153	752
00000014	22148	464
00000067	4144	54
00000302	45353	400
00000346	40000	654
00002512	33158	245

Military Mobile

Number of records included in search: 16

Keyfield	Application	Prog Lang	Dev Coun	Dev Std	Lv of Cmplx
00000034	Database	Assembly C	USA		Routine
00000303	Signal Processing	Assembly 50% PASCAL 50%	USA	Other	Complex
00000347	Other	Ada 95% Machine 5%	USA	DoD-STD-2167 (Full)	Routine
00000348	Command/Control	Ada 90% C 9% Machine 1%	USA	DoD-STD-2167 (Full)	Complex
00000349	Other	Ada 95% Machine 5%	USA	DoD-STD-2167 (Full)	Routine
00002456	Mission Planning	Ada 8% FORTRAN 92%	USA	DoD-STD-2167A (Full)	Difficult
00002483	Database	Ada 20% C 30% FORTRAN 50%	USA	Other	Complex
00002500	Signal Processing	Assembly 25% C 75%	USA		Difficult
00002502	Command/Control	Ada 95% Assembly 5%	USA	Other	Complex
00002503	MMI/Graphics	Ada 95% Assembly 5%	USA	Other	Routine
00002504	Command/Control	Ada 95% Assembly 5%	USA	Other	Complex
00002505	Command/Control	Ada 95% Assembly 5%	USA	Other	Difficult
00002506	Command/Control	Ada 95% Assembly 5%	USA	Other	Difficult
00002507	Other	Ada 95% Assembly 5%	USA	DoD-STD-1267A (Tailored)	Difficult
00002508	Command/Control	Ada 95% Assembly 5%	USA	Other	Routine
00002515	Command/Control	Ada 100%	USA	Other	Difficult

Military Mobile (cont'd)

Number of records included in search: 16

Keyfield	Norm Eff Sz	Norm Effrt
00000034	17134	83
00000303	30000	237
00000347	2311	39
00000348	18052	396
00000349	3268	56
00002456	63254	221
00002483	697814	284
00002500	1958	14
00002502	26239	633
00002503	32464	78
00002504	26239	633
00002505	7448	180
00002506	6317	152
00002507	26814	647
00002508	58789	1418
00002515	15025	13

Appendix B: Development Standard Phased Effort

		RQTS	DESIGN	CODE	TEST	total		
DOD 2167A		25%	24%	20%	31%	100%		
	w/o rqts:	0%	32%	27%	41%	100%		
DOD 483/490		23%	22%	20%	35%	100%		
	w/o rqts:	0%	29%	26%	45%	100%		
Sherry's F-2:		4.7%	26.0%	55.7%	13.7%	100.0%		
UNMANNED SPACE:			Effort by Phase					
#	Point	R	D	C	T	total calc	total given	SLOC
1	95	13.3	12.7	10.6	16.4	53	53	600
2	96	26.5	25.4	21.2	32.9	106	106	600
3	103	1.8	1.7	1.4	2.2	7	7	600
4	104	47.8	45.8	38.2	59.2	191	191	600
5	85	51.0	49.0	40.8	63.2	204	204	1950
6	92	18.8	18.0	15.0	23.3	75	75	2920
7	38	15.3	14.6	12.2	18.9	61	61	4671
8	84	199.5	191.5	159.6	247.4	798	798	6000
9	86	50.0	48.0	40.0	62.0	200	200	6000
10	83	0.0	16.3	14.8	25.9	57	57	6800
11	99	58.5	56.2	46.8	72.5	234	234	8000
12	107	40.0	38.4	32.0	49.6	160	160	8000
13	112	377.8	362.6	302.2	468.4	1511	1511	8290
14	74	0.0	22.9	20.8	36.4	80	80	11700
15	115	27.3	26.2	21.8	33.8	109	109	13000
16	76	0.0	32.9	29.9	52.3	115	115	14000
17	82	0.0	40.0	36.4	63.6	140	140	16300
18	106	51.5	49.4	41.2	63.9	206	206	16300
19	113	312.0	299.5	249.6	386.9	1248	1248	19500
20	105	1.3	1.2	1.0	1.6	5	5	21000
21	81	0.0	46.9	42.6	74.5	164	164	22900
22	78	0.0	136.6	124.2	217.3	478	478	48300
23	79	0.0	123.4	112.2	196.4	432	432	50300
24	91	292.3	280.6	233.8	362.4	1169	1169	52275
25	77	0.0	149.4	135.8	237.7	523	523	56200
26	117	163.0	156.5	130.4	202.1	652	652	66843
27	80	0.0	84.6	76.9	134.5	296	296	69450
28	3	145.8	139.9	116.6	180.7	583	583	80000
29	97	132.5	127.2	106.0	164.3	530	530	80000
30	98	21.5	20.6	17.2	26.7	86	86	90300
31	90	263.8	253.2	211.0	327.1	1055	1055	95000
32	75	0.0	260.6	236.9	414.5	912	912	116800
33	88	61.0	58.6	48.8	75.6	244	244	117000
34	114	58.8	56.4	47.0	72.9	235	235	162945
35	89	150.5	144.5	120.4	186.6	602	602	225000
36	93	100.3	96.2	80.2	124.3	401	401	250000
37	119	196.8	188.9	157.4	244.0	787	787	278488
38	118	191.3	183.6	153.0	237.2	765	765	358000
39	116	367.0	352.3	293.6	455.1	1468	1468	399635
NOTE: The validation records have been shaded in grey.								

	MIL. AVIONICS:							
			Effort by Phase					
#	Point	R	D	C	T	total calc	total given	SLOC
1	67	13.5	13.0	10.8	16.7	54	54	4144
2	12	28.0	26.9	22.4	34.7	112	112	22027
3	14	116.0	111.4	92.8	143.8	464	464	22148
4	11	49.5	47.5	39.6	61.4	198	198	32878
5	2512	61.3	58.8	49.0	76.0	245	245	33158
6	346	163.5	157.0	130.8	202.7	654	654	40000
7	10	92.5	88.8	74.0	114.7	370	370	43207
8	302	0.0	128.0	106.7	165.3	400	400	45353
9	13	188.0	180.5	150.4	233.1	752	752	58153
	MIL. MOBILE							
			Effort by Phase					
#	Point	R	D	C	T	total calc	total given	SLOC
1	2500	3.5	3.4	2.8	4.3	14	14	1958
2	347	9.8	9.4	7.8	12.1	39	39	2311
3	349	14.0	13.4	11.2	17.4	56	56	3268
4	2506	7.1	39.5	84.6	20.8	152	152	6317
5	2505	8.4	46.7	100.2	24.7	180	180	7448
6	2515	0.6	3.4	7.2	1.8	13	13	15025
7	34	20.8	19.9	16.6	25.7	83	83	17134
8	348	99.0	95.0	79.2	122.8	396	396	18052
9	2502	29.6	164.3	352.3	86.7	633	633	26239
10	2507	161.8	155.3	129.4	200.6	647	647	26814
11	303	11.1	61.5	131.9	32.5	237	237	30000
12	2503	3.7	20.2	43.4	10.7	78	78	32464
13	2508	66.4	368.1	789.3	194.3	1418	1418	58789
14	2456	55.3	53.0	44.2	68.5	221	221	63254
15	2483	13.3	73.7	158.1	38.9	284	284	697814

#	Record #	R	D	C	T	total calc	total given	SLOC
	MIL. GROUND							
			Effort by Phase					
1	1	2.3	2.2	1.8	2.8	9	9	1500
2	26	5.3	5.0	4.2	6.5	21	21	4170
3	23	10.0	9.6	8.0	12.4	40	40	5200
4	2526	0.0	67.5	56.3	87.2	211	211	6681
5	2527	0.0	75.2	62.7	97.1	235	235	7457
6	155	18.5	17.8	14.8	22.9	74	74	8398
7	154	27.3	26.2	21.8	33.8	109	109	8965
8	2612	0.0	99.8	83.2	129.0	312	312	9899
9	2497	3.6	19.7	42.3	10.4	76	76	10000
10	153	37.3	35.8	29.8	46.2	149	149	11534
11	63	5.1	4.8	4.4	7.7	22	22	11753
12	2611	0.0	119.7	99.7	154.6	374	374	11840
13	136	38.5	37.0	30.8	47.7	154	154	12121
14	138	47.5	45.6	38.0	58.9	190	190	14389
15	2610	0.0	146.6	122.1	189.3	458	458	14536
16	59	32.0	30.6	27.8	48.7	139	139	15554
17	5	12.3	11.8	9.8	15.2	49	49	16000
18	127	3.3	3.1	2.6	4.0	13	13	16016
19	39	0.0	6.1	5.1	7.9	19	19	17000
20	24	56.5	54.2	45.2	70.1	226	226	18000
21	145	25.3	24.2	20.2	31.3	101	101	18560
22	53	11.0	10.6	9.6	16.8	48	48	21122
23	2528	0.0	217.9	181.6	281.5	681	681	21588
24	150	25.0	24.0	20.0	31.0	100	100	21681
25	56	5.1	4.8	4.4	7.7	22	22	22150
26	129	11.0	10.6	8.8	13.6	44	44	22516
27	40	13.5	13.0	10.8	16.7	54	54	23483
28	143	21.5	20.6	17.2	26.7	86	86	23703
29	135	66.0	63.4	52.8	81.8	264	264	23787
30	124	34.8	33.4	27.8	43.1	139	139	23881
31	72	11.3	10.8	9.0	14.0	45	45	25160
32	120	23.8	22.8	19.0	29.5	95	95	25842
33	58	139.4	133.3	121.2	212.1	606	606	28191
34	142	87.0	83.5	69.6	107.9	348	348	28782
35	2	58.0	55.7	46.4	71.9	232	232	28805
36	131	48.0	46.1	38.4	59.5	192	192	29147
37	144	36.3	34.8	29.0	45.0	145	145	29802
38	147	48.0	46.1	38.4	59.5	192	192	31720
39	151	47.5	45.6	38.0	58.9	190	190	38174
40	139	59.3	56.9	47.4	73.5	237	237	38634
41	6	16.5	15.8	13.2	20.5	66	66	41000
42	30	38.0	36.5	30.4	47.1	152	152	41000
43	134	57.0	54.7	45.6	70.7	228	228	44527
44	54	29.2	27.9	25.4	44.5	127	127	45035
45	7	30.0	28.8	24.0	37.2	120	120	45057
46	132	69.5	66.7	55.6	86.2	278	278	46595

47	123	44.0	42.2	35.2	54.6	176	176	47600
48	126	41.3	39.6	33.0	51.2	165	165	47965
49	137	68.5	65.8	54.8	84.9	274	274	60233
50	152	71.5	68.6	57.2	88.7	286	286	69772
51	140	1.5	1.4	1.2	1.9	6	6	70020
52	130	184.5	177.1	147.6	228.8	738	738	71851
53	61	24.6	23.5	21.4	37.5	107	107	72716
54	301	93.8	89.8	81.6	142.8	408	408	76566
55	2517	7.8	43.3	93.0	22.9	167	167	85382
56	2521	156.5	150.2	125.2	194.1	626	626	97087
57	2498	0.7	4.2	8.9	2.2	16	16	100000
58	2501	16.7	92.4	198.1	48.8	356	356	110400
59	25	180.0	172.8	144.0	223.2	720	720	111995
60	28	210.3	201.8	168.2	260.7	841	841	112917
61	133	161.3	154.8	129.0	200.0	645	645	123710
62	9	129.3	124.1	103.4	160.3	517	517	128200
63	48	81.5	78.2	65.2	101.1	326	326	136227
64	50	0.0	218.9	182.4	282.7	684	684	144000
65	141	80.5	77.3	64.4	99.8	322	322	162039
66	2525	248.3	238.3	198.6	307.8	993	993	200000
67	2523	237.3	227.8	189.8	294.2	949	949	231018
68	55	1463.5	1399.9	1272.6	2227.1	6363	6363	260882
69	148	637.8	612.2	510.2	790.8	2551	2551	300000
70	2524	324.5	311.5	259.6	402.4	1298	1298	363371
71	2519	390.0	374.4	312.0	483.6	1560	1560	419619
72	2520	842.5	808.8	674.0	1044.7	3370	3370	419619
73	2522	547.8	525.8	438.2	679.2	2191	2191	461426

Appendix C: Productivity Calibration Constants

Mil Ground						
	(1.2xApp)		(.87xApp)		Class	
Iteration	System	Application	Support	Reference	Multiplier	Correlation
Linear	1.3724	1.1437	0.9950	1.1437	1.0000	0.931
Space						
	(1.2xApp)		(.87xApp)		Class	
Iteration	System	Application	Support	Reference	Multiplier	Correlation
Linear	0.2570	0.2142	0.1864	1.1437	0.1873	0.324
Log	4.7705	3.9754	3.4586	1.1437	3.4759	
Avionics						
	(1.2xApp)		(.87xApp)		Class	
Iteration	System	Application	Support	Reference	Multiplier	Correlation
Linear	2.0804	1.7337	1.5083	1.1437	1.5159	0.777
Mil Mobile						
	(1.2xApp)		(.87xApp)		Class	
Iteration	System	Application	Support	Reference	Multiplier	Correlation
Linear	2.2390	1.8658	1.6232	1.1437	1.6314	0.643

Appendix D: Estimates and Statistics

MIL. GROUND:							
1. ***Uncalibrated*** validation points:							
							%
#	Record	**Saset**	Actual	Diff	% diff	MRE	Counter
3	23	59	40	-19	-48%	0.48	0
6	155	95	74	-21	-28%	0.28	0
9	2497	113	76	-37	-49%	0.49	0
12	2611	134	374	240	64%	0.64	0
15	2610	164	458	294	64%	0.64	0
18	127	181	13	-168	-1292%	12.92	0
21	145	210	101	-109	-108%	1.08	0
24	150	245	100	-145	-145%	1.45	0
27	40	266	54	-212	-393%	3.93	0
30	124	270	139	-131	-94%	0.94	0
33	58	319	606	287	47%	0.47	0
36	131	330	192	-138	-72%	0.72	0
39	151	432	190	-242	-127%	1.27	0
42	30	464	152	-312	-205%	2.05	0
45	7	510	120	-390	-325%	3.25	0
48	126	543	165	-378	-229%	2.29	0
51	140	792	6	-786	-13100%	131.00	0
54	301	866	408	-458	-112%	1.12	0
57	2498	1131	16	-1115	-6969%	69.69	0
60	28	1277	841	-436	-52%	0.52	0
63	48	1541	326	-1215	-373%	3.73	0
66	2525	2262	993	-1269	-128%	1.28	0
69	148	3393	2551	-842	-33%	0.33	0
72	2520	4746	3370	-1376	-41%	0.41	0
						240.98	0
Mean				-374.08			
Var				235349			
Std D				485.128			
MMRE				10.04			
% Desc				0%			

NOTE: The "% Counter" column is used in the computation of the "% Desc" statistic. Also, th
"Diff" column consists of Actual - Estimate.

2. ***Calibrated*** validation points:

#	Record	Linear	Actual	Diff	% diff	MRE	% Counter
3	23	35	40	5	13%	0.13	1
6	155	57	74	17	23%	0.23	1
9	2497	68	76	8	11%	0.11	1
12	2611	81	374	293	78%	0.78	0
15	2610	99	458	359	78%	0.78	0
18	127	109	13	-96	-738%	7.38	0
21	145	126	101	-25	-25%	0.25	1
24	150	148	100	-48	-48%	0.48	0
27	40	160	54	-106	-196%	1.96	0
30	124	163	139	-24	-17%	0.17	1
33	58	192	606	414	68%	0.68	0
36	131	198	192	-6	-3%	0.03	1
39	151	260	190	-70	-37%	0.37	0
42	30	279	152	-127	-84%	0.84	0
45	7	307	120	-187	-156%	1.56	0
48	126	327	165	-162	-98%	0.98	0
51	140	477	6	-471	-7850%	78.50	0
54	301	521	408	-113	-28%	0.28	0
57	2498	681	16	-665	-4156%	41.56	0
60	28	769	841	72	9%	0.09	1
63	48	928	326	-602	-185%	1.85	0
66	2525	1362	993	-369	-37%	0.37	0
69	148	2043	2551	508	20%	0.20	1
72	2520	2857	3370	513	15%	0.15	1
						139.73	9
Mean				-36.75			
Var				94006.2			
Std D				306.604			
MMRE				5.82			
% Desc				38%			

UNMANNED SPACE:

1. ***Uncalibrated*** validation points:

#	Record	**Saset**	Actual	Diff	% diff	MRE	% Counter
3	103	16	7	-9	-129%	1.29	0
6	92	76	75	-1	-1%	0.01	1
9	86	156	200	44	22%	0.22	1
12	107	208	160	-48	-30%	0.30	0
15	115	338	109	-229	-210%	2.10	0
18	106	424	206	-218	-106%	1.06	0
21	81	596	164	-432	-263%	2.63	0
24	91	1360	1169	-191	-16%	0.16	1
27	80	1807	296	-1511	-510%	5.10	0
30	98	2349	86	-2263	-2631%	26.31	0
33	88	3044	244	-2800	-1148%	11.48	0
36	93	6504	401	-6103	-1522%	15.22	0
39	116	10396	1468	-8928	-608%	6.08	0
						71.97	3
Mean				-1745.3			
Var				7744016			
Std D				2782.81			
MMRE				5.54			
% Desc				23%			

2. ***Calibrated*** validation points:

#	Record	**Linear**	Actual	Diff	% diff	MRE	% Counter
3	103	0.1	7	6.9	99%	0.99	0
6	92	0.7	75	74.3	99%	0.99	0
9	86	1.4	200	198.6	99%	0.99	0
12	107	1.9	160	158.1	99%	0.99	0
15	115	3.1	109	105.9	97%	0.97	0
18	106	3.9	206	202.1	98%	0.98	0
21	81	5.5	164	158.5	97%	0.97	0
24	91	12.5	1169	1156.5	99%	0.99	0
27	80	16.6	296	279.4	94%	0.94	0
30	98	21.6	86	64.4	75%	0.75	0
33	88	27.9	244	216.1	89%	0.89	0
36	93	59.7	401	341.3	85%	0.85	0
39	116	95.4	1468	1372.6	94%	0.94	0
						12.23	0
Mean				333.438			
Var				180729			
Std D				425.123			
MMRE				0.94			
% Desc				0%			

UNMANNED SPACE (continued):

#	Record	Log	Actual	Diff	% diff	MRE	% Counter
3	103	49	7	-42	-600%	6.00	0
6	92	240	75	-165	-220%	2.20	0
9	86	494	200	-294	-147%	1.47	0
12	107	658	160	-498	-311%	3.11	0
15	115	1069	109	-960	-881%	8.81	0
18	106	1341	206	-1135	-551%	5.51	0
21	81	1884	164	-1720	-1049%	10.49	0
24	91	4300	1169	-3131	-268%	2.68	0
27	80	5713	296	-5417	-1830%	18.30	0
30	98	7428	86	-7342	-8537%	85.37	0
33	88	9624	244	-9380	-3844%	38.44	0
36	93	20565	401	-20164	-5028%	50.28	0
39	116	32873	1468	-31405	-2139%	21.39	0
						254.06	0
Mean				-6281			
Var				8.9E+07			
Std D				9427.65			
MMRE				19.54			
% Desc				0%			

MILITARY AVIONICS:									
1. ***Uncalibrated*** validation points:									
							%		
#	Record	**Saset**	Actual	Diff	% diff	MRE	Counter		
5	2512	675	245	-430	-176%	1.76	0		
						1.76	0		
Mean				-430					
Var				N/A					
Std D				N/A					
MMRE				1.76					
% Desc				0%					
2. ***Calibrated*** validation points:									
							%		
#	Record	**Linear**	Actual	Diff	% diff	MRE	Counter		
5	2512	300	245	-55	-22%	0.22	1		
						0.22	1		
Mean				-55					
Var				N/A					
Std D				N/A					
MMRE				0.22					
% Desc				100%					

MILITARY MOBILE:

1. ***Uncalibrated*** validation points:

#	Record	Saset	Actual	Diff	% diff	MRE	% Counter
3	349	50	56	6	11%	0.11	1
6	2515	229	13	-216	-1662%	16.62	0
9	2502	401	633	232	37%	0.37	0
12	2503	496	78	-418	-536%	5.36	0
						22.45	1
Mean				-99			
Var				78678.7			
Std D				280.497			
MMRE				5.61			
% Desc				25%			

2. ***Calibrated*** validation points:

#	Record	Linear	Actual	Diff	% diff	MRE	% Counter
3	349	32	56	24	43%	0.43	0
6	2515	146	13	-133	-1023%	10.23	0
9	2502	255	633	378	60%	0.60	0
12	2503	315	78	-237	-304%	3.04	0
						14.29	0
Mean				8			
Var				72354			
Std D				268.987			
MMRE				3.57			
% Desc				0%			

Appendix E: Wilcoxon Test

MIL. GROUND:

1. ***Uncalibrated*** validation points:

#	Record	Saset	Actual	Diff	Rank +	Rank -
3	23	59	40	-19		1
6	155	95	74	-21		2
9	2497	113	76	-37		3
12	2611	134	374	240	10	
15	2610	164	458	294	13	
18	127	181	13	-168		8
21	145	210	101	-109		4
24	150	245	100	-145		7
27	40	266	54	-212		9
30	124	270	139	-131		5
33	58	319	606	287	12	
36	131	330	192	-138		6
39	151	432	190	-242		11
42	30	464	152	-312		14
45	7	510	120	-390		16
48	126	543	165	-378		15
51	140	792	6	-786		19
54	301	866	408	-458		18
57	2498	1131	16	-1115		21
60	28	1277	841	-436		17
63	48	1541	326	-1215		22
66	2525	2262	993	-1269		23
69	148	3393	2551	-842		20
72	2520	4746	3370	-1376		24
				Total	35	265

Wilcoxon T = 35

n = 24

P	To	Result
0.10	92	Reject
0.05	81	Reject

Since Rank + ≤ To, SASET appears to consistently overestimate.

NOTE: n is the number of observations, P is the alpha level of confidence, To is the critical value, and Result refers to whether T ≤ To (Reject null hypothesis) or T ≥ To (Accept null hypothesis). The null hypothesis is that the two population distributions are the same.

2. ***Calibrated*** validation points:

#	Record	Linear	Actual	Diff	Rank +	Rank -
3	23	35	40	5	1	
6	155	57	74	17	4	
9	2497	68	76	8	3	
12	2611	81	374	293	16	
15	2610	99	458	359	17	
18	127	109	13	-96		10
21	145	126	101	-25		6
24	150	148	100	-48		7
27	40	160	54	-106		11
30	124	163	139	-24		5
33	58	192	606	414	19	
36	131	198	192	-6		2
39	151	260	190	-70		8
42	30	279	152	-127		13
45	7	307	120	-187		15
48	126	327	165	-162		14
51	140	477	6	-471		20
54	301	521	408	-113		12
57	2498	681	16	-665		24
60	28	769	841	72	9	
63	48	928	326	-602		23
66	2525	1362	993	-369		18
69	148	2043	2551	508	21	
72	2520	2857	3370	513	22	
				Total	112	188

Wilcoxon T = 112

n = 24

P	To	Result
0.10	92	Accept
0.05	81	Accept

We can not conclude that the distributions are not the same. Also, neither Rank + nor Rank - are $\leq$ To; indicating that there appears to be no bias.

UNMANNED SPACE:

1. ***Uncalibrated*** validation points:

#	Record	**Saset**	Actual	Diff	Rank +	Rank -
3	103	16	7	-9		2
6	92	76	75	-1		1
9	86	156	200	44	3	
12	107	208	160	-48		4
15	115	338	109	-229		7
18	106	424	206	-218		6
21	81	596	164	-432		8
24	91	1360	1169	-191		5
27	80	1807	296	-1511		9
30	98	2349	86	-2263		10
33	88	3044	244	-2800		11
36	93	6504	401	-6103		12
39	116	10396	1468	-8928		13
				Total	3	88

Wilcoxon T = 3

n = 13

P	To	Result
0.10	21	Reject
0.05	17	Reject

Since Rank + ≤ To, SASET appears to consistently overestimate.

2. ***Calibrated*** validation points:

#	Record	**Linear**	Actual	Diff	Rank +	Rank -
3	103	0.1	7	6.9	1	
6	92	0.7	75	74.3	3	
9	86	1.4	200	198.6	7	
12	107	1.9	160	158.1	5	
15	115	3.1	109	105.9	4	
18	106	3.9	206	202.1	8	
21	81	5.5	164	158.5	6	
24	91	12.5	1169	1156.5	12	
27	80	16.6	296	279.4	10	
30	98	21.6	86	64.4	2	
33	88	27.9	244	216.1	9	
36	93	59.7	401	341.3	11	
39	116	95.4	1468	1372.6	13	
				Total	91	-

Wilcoxon T = 91

n = 13

P	To	Result
0.10	21	Accept
0.05	17	Accept

We can not conclude that the distributions are not the same. However, since Rank - ≤ To, we can say that the model appears to consistently underestimate.

UNMANNED SPACE (continued):

#	Record	Log	Actual	Diff	Rank +	Rank -
3	103	49	7	-42		1
6	92	240	75	-165		2
9	86	494	200	-294		3
12	107	658	160	-498		4
15	115	1069	109	-960		5
18	106	1341	206	-1135		6
21	81	1884	164	-1720		7
24	91	4300	1169	-3131		8
27	80	5713	296	-5417		9
30	98	7428	86	-7342		10
33	88	9624	244	-9380		11
36	93	20565	401	-20164		12
39	116	32873	1468	-31405		13
				Total	-	91

Wilcoxon T = 91

n = 13

P	To	Result
0.10	21	Accept
0.05	17	Accept

We can not conclude that the distributions are not the same. However, since Rank + ≤ To, we can say that the model appears to consistently overestimate.

MILITARY AVIONICS:

1. ***Uncalibrated*** validation points:

#	Record	**Saset**	Actual	Diff
5	2512	675	245	-430

The minimum number of data points represented in the To table are n = 5 (Mendenhall, 1990)

2. ***Calibrated*** validation points:

#	Record	**Linear**	Actual	Diff
5	2512	300	245	-55

The minimum number of data points represented in the To table are n = 5 (Mendenhall, 1990)

MILITARY MOBILE:

1. ***Uncalibrated*** validation points:

#	Record	**Saset**	Actual	Diff
3	349	50	56	6
6	2515	229	13	-216
9	2502	401	633	232
12	2503	496	78	-418

The minimum number of data points represented in the To table are n = 5 (Mendenhall, 1990)

2. ***Calibrated*** validation points:

#	Record	**Linear**	Actual	Diff
3	349	32	56	24
6	2515	146	13	-133
9	2502	255	633	378
12	2503	315	78	-237

The minimum number of data points represented in the To table are n = 5 (Mendenhall, 1990)

Appendix F: SAS Output

Plot of EFFORT*SLOC. Legend: A = 1 obs, B = 2 obs, etc.

```
EFFORT |
  7000 +
       |
       |
       |
       |                    A
       |
  6000 +
       |
       |
       |
       |
       |
  5000 +
       |
       |
       |
       |
  4000 +
       |
       |
       |
       |                               A
       |
  3000 +
       |
       |
       |                      A
       |
       |                                  A
  2000 +
       |
       |                               A
       |
       |
  1000 +                  A  A
       |        A
       | A    A    A A
       | A
       |ABAB  A    A
       |CEDAB  A
     0 +BBDB A   A
       |
       -+---------+---------+---------+---------+---------+
        0      100000    200000    300000    400000    500000

                          SLOC
```

The SAS System

Model: MODEL1
Dependent Variable: EFFORT

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	29034513.92	29034513.92	50.474	0.0001
Error	47	27035892.08	575231.74639		
C Total	48	56070406			

Root MSE	758.44034	R-square	0.5178
Dep Mean	562.57143	Adj R-sq	0.5076
C.V.	134.81672		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0
INTERCEP	1	0.794511	134.13410265	0.006
SLOC	1	0.006827	0.00096097	7.105

Variable	DF	Prob > |T|
INTERCEP	1	0.9953
SLOC	1	0.0001

EFFORT
The SAS System

Model: MODEL2
Dependent Variable: EFFORT

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	27453212.497	27453212.497	45.088	0.0001
Error	47	28617193.503	608876.45751		
C Total	48	56070406			

Root MSE	780.30536	R-square	0.4896
Dep Mean	562.57143	Adj R-sq	0.4788
C.V.	138.70334		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0
INTERCEP	1	-551.731625	199.91173178	-2.760
SQRTSLOC	1	4.679635	0.69691529	6.715

Variable	DF	Prob > |T|
INTERCEP	1	0.0082
SQRTSLOC	1	0.0001

The SAS System

Model: MODEL3
Dependent Variable: EFFORT

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	28875432.675	28875432.675	49.904	0.0001
Error	47	27194973.325	578616.45373		
C Total	48	56070406			

Root MSE	760.66843	R-square	0.5150

Dep Mean	562.57143	Adj R-sq	0.5047
C.V.	135.21277		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0
INTERCEP	1	-184.530272	151.63486347	-1.217
QSLOC	1	0.175883	0.02489741	7.064

Variable	DF	Prob > |T|
INTERCEP	1	0.2297
QSLOC	1	0.0001

The SAS System

Model: MODEL4
Dependent Variable: EFFORT

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	28556941.488	28556941.488	48.783	0.0001
Error	47	27513464.512	585392.86196		
C Total	48	56070406			

Root MSE	765.10971	R-square	0.5093
Dep Mean	562.57143	Adj R-sq	0.4989
C.V.	136.00223		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0
INTERCEP	1	-277.879386	162.56235368	-1.709
HEPTSLOC	1	0.526168	0.07533416	6.984

Variable	DF	Prob > |T|
INTERCEP	1	0.0940
HEPTSLOC	1	0.0001

The SAS System

Model: MODEL5
Dependent Variable: EFFORT

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	19860823.109	19860823.109	25.779	0.0001
Error	47	36209582.891	770416.65726		
C Total	48	56070406			

Root MSE	877.73382	R-square	0.3542
Dep Mean	562.57143	Adj R-sq	0.3405
C.V.	156.02176		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0
INTERCEP	1	-4745.234728	1052.8854095	-4.507
LNSLOC	1	503.430069	99.15242997	5.077

Variable	DF	Prob > |T|
INTERCEP	1	0.0001
LNSLOC	1	0.0001

REFERENCES

Boehm, B.W. Software Engineering Economics. Englewood Cliffs NJ: Prentice-Hall, Inc., 1981.

-----. "Software Engineering Economics," IEEE Transactions on Software Engineering, 1: 239-256 (1984).

Bowden, R.G., Cheadle, W.G., & Ratliff, R.W. SASET 3.0 Technical Reference Manual. Publication S-3730-93-2. Denver: Martin Marietta Astronautics Group, 1993.

-----. SASET 3.0 User's Guide. Publication S-3730-93-1. Denver: Martin Marietta Astronautics Group, 1993.

Brooks, F.P. Jr. The Mythical Man-Month: Essays on Software Engineering. Menlo Park CA: Addison-Wesley, 1975.

Charette, R.N. Software Engineering Environments: Concepts and Technology. New York: Intertext Publications, Inc., 1986.

Cheadle, W.G., Herrington, J.L., Mogensen, C.H., Suhr, J.D. Software Cost Estimation Study: SASET Baseline Model (Revision Six). Requirements Document. Denver: Martin Marietta Astronautics Group, July 1988.

Coggins, G.A., & Russell, R.C. Software Cost Estimating Models: A Comparative Study of What the Models Estimate. MS thesis, AFIT/GCA/LAS/93S-4. School of Systems and Logistics, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, 1993 (AD-A275989).

Conte, S.D., Dunsmore H.E., Shen V.Y. Software Engineering Metrics and Models. Menlo Park CA: The Benjamin/Cummings Publishing Company, Inc., 1986.

Daly, B.A. A Comparison of Software Schedule Estimators. MS thesis, AFIT/GCA/LSQ/90S-1. School of Systems and Logistics, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, 1990 (AD-A229532).

Devore, J.L. Probability and Statistics for Engineering and the Sciences. Belmont CA: Duxbury Press, 1991.

Ferens, D. Class handout, COST 677, Quantitative Management of Software. School of Systems and Logistics, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, Fall Quarter 1994.

Fulton, R. & Stukes, S. SMC SWDB User's Manual: Version 1.0. Oxnard CA: Management Consulting & Research, Inc., 1995.

Harbert, C.E., & Ratliff, R.W. Database Management System and Calibration Tool: DBMS Version 1.4 User's Guide. Publication S-3730-93-3. Denver: Martin Marietta Astronautics Group, 1993.

IITRI. "Test Case Study: Estimating the Cost of Ada Software Development," Lanham MD, 1989.

Maness, R. Software Engineer, Lockheed Martin, NJ. Telephone interview. 19 April 1995 and 25 April 1995.

Mendenhall, W., Wackerly, D, & Scheaffer, R. Mathematical Statistics with Applications (Fourth Edition). Belmont CA: Duxbury Press, 1990.

Neter, J., Wasserman, W., & Kutner, M. Applied Linear Regression Models (Second Edition). Burr Ridge IL: Irwin, 1989.

Ourada, G. L. Software Cost Estimating Models: A Calibration, Validation, and Comparison. MS thesis AFIT/GSS/LSY/91D-11. School of Systems and Logistics, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, 1991 (AD-A246677).

Pighetti, T. Software Engineer, Lockheed Martin, Denver CO. Telephone interview. 17 February 1995 and 24 April 1995.

Silver, A.N., & Cheadle, W.G. Software Cost Estimation Study: Cost Drivers Report. Technical Report, Contract N00014-85-C-0892. Denver: Martin Marietta Aerospace Corporation, June 1986.

-----. Software Cost Estimation Study: CER Methodology Prototype. Technical Report, Contract N00014-85-C-0892. Denver: Martin Marietta Aerospace Corporation, October 1986.

-----. Software Cost Estimation Study: CER Model Planning Report. Technical Report, Contract N00014-85-C-0892. Denver: Martin Marietta Aerospace Corporation, April 1987.

Stukes, S., & Apgar, H. Air Force Cost Analysis Agency Software Model Content Study. Final report TR-9359/51-8. Oxnard CA: Management Consulting & Research, Inc., 1994a.

Stukes, S., Apgar, H., Galorath, D., & Maness, R. Application Oriented Software Data Collection: Software Model Calibration Report. TR-9007/49-1. Oxnard CA: Management Consulting & Research, Inc., 1991b.

Thibodeau, R. "An Evaluation of Software Cost Estimating Models." New York: Rome Air Development Center, 1981.

Wellman, F. Software Costing: An Objective Approach to Estimating and Controlling the Cost of Computer Software. New York: Prentice-Hall, Inc., 1992.

Vita

On May 27, 1992 Carl "Dave" Vegas graduated as a Second Lieutenant from the United States Air Force Academy with a Bachelor of Science degree and a major in management. His first assignment was a brief tour at Homestead AFB, Florida, which was closed by Hurricane Andrew in August, 1992. Next he was stationed at MacDill AFB, Florida where he worked as a financial analyst. He attended the Financial Analysis Officer Course at Sheppard AFB, Texas in the spring of 1993. Six months after returning from the course he was promoted to Deputy Chief of the Financial Analysis Office of the 56 Comptroller Squadron. In May, 1994 he arrived at the Air Force Institute of Technology at Wright-Patterson AFB, Ohio as a graduate student. A few days after arriving he was promoted to the rank of First Lieutenant. On September 26, 1995 he graduated with a Master in Cost Analysis. His follow on assignment was McClellan AFB, California.

Permanent Address: 15448 S.W. 148 St.

Miami, FL 33196

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
September 1995

3. REPORT TYPE AND DATES COVERED
Master's Thesis

4. TITLE AND SUBTITLE
CALIBRATION OF THE SOFTWARE ARCHITECTURE SIZING AND ESTIMATION TOOL (SASET)

5. FUNDING NUMBERS

6. AUTHOR(S)
Carl D. Vegas, 1st Lieutenant, USAF

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Institute of Technology,
WPAFB OH 45433-6583

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT/GCA/LAS/95S-11

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
USAF SMC
El Segundo, CA 90245-4687

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This study attempted to analyze the effect of calibration on the performance of the SASET computer software cost estimating model. Data used for input into the model were drawn from the most current USAF SMC Software Database (SWDB). Once all the records to be used for analysis were identified, the DBMS/Calibration tool (which is part of SASET) was used to perform regression analysis on the relationship between program size (measured in SLOC) and the effort required to develop the program (measured in man-months). Productivity information reported from this tool was then input into equations used to calculate the Productivity Calibration Constants (PCC) and Software Class Multipliers. A comparison was then made between the model's accuracy before calibration and its accuracy after calibration. This was done using records which were not used in calibration (referred to as validation points). Several measures such as mean, variance, mean magnitude of relative error (MMRE), and the percentage method were used to describe accuracy. The majority of the results agreed with previous studies that calibration does improve a model's prediction performance. However, emphasis is placed on the fact that calibration is most useful when the group of calibration data points are homogenous.

14. SUBJECT TERMS
Calibration, Software, Cost Estimation, Cost Model, Validation, Regression, SASET, Parametric Analysis, DBMS, Space Projects, Accuracy.

15. NUMBER OF PAGES
106

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL